

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

run on: June 8, 2004, 13:55:43 ; Search time 91.7861 Seconds
(without alignments)
1065.101 Million cell updates/sec

title: US-10-010-050A-2

perfect score: 1927

sequence: 1 MRGAGARGASWCWALAL.....IKITYEIPLPFRKKTISGL 346

scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

sarched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

post-processing: Minimum March 0%
Maximum March 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1920	99.6	346	2	AAW92967 Human zsi
2	1916	99.4	347	2	AAV41323 Human sec
3	1913	99.3	358	4	AAW93870 Human pol
4	886	46.0	163	4	AAW19604 Peptide #
5	886	46.0	163	4	ABH39239 Peptide #
6	886	46.0	163	4	AAW32740 Peptide #
7	886	46.0	163	4	ABH24080 Protein #
8	886	46.0	163	4	AAW72492 Human bon
9	886	46.0	163	4	AAW59901 Human bra
10	886	46.0	163	4	ABG54183 Human liv
11	886	46.0	163	5	ABG42312 Human pep
12	472	24.5	89	5	ABP42902 Human ova
13	96.5	5.0	1705	5	ABR52348 Protein r
14	95.5	5.0	2703	4	ABH60074 Drosophil
15	94	4.9	429	5	ABH52308 Helicobac
16	94	4.9	1238	6	ABH30825 Protein e
17	93	4.8	478	6	ABU27691 Protein e
18	92.5	4.8	569	4	ABH71458 Drosophil
19	91	4.7	433	3	AAW31636 Arabidops
20	91	4.7	440	2	AAW41852 Ragweed p
21	91	4.7	508	3	AAW31635 Arabidops
22	91	4.7	742	3	AAW94977 Human sec
23	91	4.7	742	7	ABH56938 Human pro
24	91	4.7	3923	2	AAW31237 Human Apo
25	91	4.7	4563	6	ABU79140 Apolipop

26	91	4.7	4590	4	AAU33184 Novel hum
27	89.5	4.6	442	2	AAW41856 Ragweed p
28	89.5	4.6	2237	5	ABG70004 Larval vi
29	89	4.6	568	4	ABH60986 Drosophil
30	88.5	4.6	1462	7	ADD14088 Human src
31	88.5	4.6	1468	4	AAW39218 Human pol
32	88.5	4.6	1496	4	AAW41004 Human pol
33	88	4.6	605	6	ABU32174 Protein e
34	88	4.6	3194	6	ABU31122 Protein e
35	87.5	4.5	311	4	AAE06730 Human CAS
36	87	4.5	226	3	ABH34827 Gene 17 h
37	87	4.5	502	2	AAW58861 T. haloph
38	87	4.5	661	2	AAW15565 Human int
39	86.5	4.5	388	5	ABH92540 Helicida
40	86	4.5	411	5	ABH49050 listeria
41	86	4.5	411	6	ABU32653 Protein e
42	85.5	4.4	428	5	ABU74922 Synthetic
43	85.5	4.4	5035	2	AAW25450 MH mutant
44	84.5	4.4	1462	2	AAW37508 Human DNA
45	83.5	4.3	373	3	AAW52581 Helicobac

ALIGNMENTS

RESULT 1	AAW92967	standard; protein; 346 AA.
AAW92967	14-MAY-1999	(first entry)
XX	AAW92967;	
AC	14-MAY-1999	(first entry)
DT	14-MAY-1999	(first entry)
XX	Human zsig46 protein.	
DE	Human zsig46 protein.	
XX	Secreted protein; zsig46; human; chromosome 13; thyroid; disease;	
XX	hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;	
KW	Hirschprung's disease; neuronal ceroid-lipofusiosis; Wilson disease;	
KW	Reiger syndrome; immunosassay; detection; anti-idiotypic antibody;	
XX	therapy; diagnostic.	
OS	Homo sapiens.	
XX	W09905275-A1.	
PN	04-FEB-1999.	
PD	04-FEB-1999.	
XX	24-JUL-1998;	98WO-US015431.
XX	24-JUL-1997;	97US-0053613P.
PR	(ZYMO) ZYMOGENETICS INC.	
PA	Sheppard PO, Gilbertson DG;	
PI	WPI; 1999-142930/12.	
DR	N-PSDB; AAW02855.	
XX	New secreted polypeptide, zsig46, and its fragments, related fusion	
PT	proteins - used for diagnosis and treatment of thyroid disorders or	
PT	diseases involving genes on chromosome 13.	
XX	Claim 3; Page 90-91; 101pp; English.	
PS	This invention describes the isolation of a novel human secreted protein,	
XX	zsig46 encoded by a gene on chromosome 13 which is mainly expressed in	
CC	the thyroid. This product can be used to study secretion of proteins from	
CC	cells and also to treat or prevent deficient expression of zsig46, which	
CC	may be associated with thyroid diseases (e.g. hypothyroidism, Graves'	
CC	disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that	
CC	involve genes in the same region of chromosome 13 (e.g. Hirschprung's	
CC	disease, neuronal ceroid-lipofusiosis, Wilson disease and Reiger	
CC	syndrome). Antibodies and other binding proteins, are used as immunoassay	

```

1C reagents to detect zsig46 or cells expressing it, e.g. for assessing
1C thyroid function to produce anti-idiotypic antibodies, for affinity
1C purification of zsig46, to screen expression libraries, to neutralise
1C zsig46 activity, and to deliver toxins, radioisotopes etc. for
1C therapeutic or diagnostic purposes. Agonists of the product can be used
1C to promote growth, differentiation and proliferation of specific cell
1C types, e.g. for treating (extra)thyroid diseases or as additive to cell
1C cultures
1C
1C Sequence 346 AA:
1C
1C Query Match 99.6%; Score 1920; DB 2; Length 346;
1C Best Local Similarity 99.7%; Pred. No. 4.2e-185;
1C Matches 345; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1C
1C 1 MRGAGAGARGASGCMALALMLAVPGMSRVSGIPSRHMPVYKRPDRPKDPYCOA 6C
1C 1 MGRGAGARGASGCMALALMLAVPGMSRVSGIPSRHMPVYKRPDRPKDPYCOA 6C
1C
1C 61 KYTFCPTGSPIPVWEGDDIEVFRLQAPVWFEFKYGDLLGHKIMHDAIGFRSTLTGKNT 120
1C 61 KYTFCPTGSPIPVWEGDDIEVFRLQAPVWFEFKYGDLLGHKIMHDAIGFRSTLTGKNT 120
1C
1C 121 MEWYELFQLGNCITFPHLRPEMDAPFMCNOGAACFEIGIDVHWKENGTLVQVATISGMF 180
1C 121 MEWYELFQLGNCITFPHLRPEMDAPFMCNOGAACFEIGIDVHWKENGTLVQVATISGMF 180
1C
1C 181 NQAKWKQKQNEGTIYETNNVKAPEKGAETWDSYDSCSFVLRRTNKALEFAERKNI 240
1C 181 NQAKWKQKQNEGTIYETNNVKAPEKGAETWDSYDSCSFVLRRTNKALEFAERKNI 240
1C
1C 241 ETNTTRIFLYSGEPTVGNESVGPCKNTGLAIRFYPPPHIPTKEFLLSLQIF 300
1C 241 ETNTTRIFLYSGEPTVGNESVGPCKNTGLAIRFYPPPHIPTKEFLLSLQIF 300
1C
1C 241 ETNTTRIFLYSGEPTVGNESVGPCKNTGLAIRFYPPPHIPTKEFLLSLQIF 300
1C 241 ETNTTRIFLYSGEPTVGNESVGPCKNTGLAIRFYPPPHIPTKEFLLSLQIF 300
1C
1C 301 DAVIVHKQFYLFYNEFWFLPMKPPFIKITYEELPLIRNKTLISGL 346
1C 301 DAVIVHKQFYLFYNEFWFLPMKPPFIKITYEELPLIRNKTLISGL 346
1C
1C 301 DAVIVHKQFYLFYNEFWFLPMKPPFIKITYEELPLIRNKTLISGL 346
1C 301 DAVIVHKQFYLFYNEFWFLPMKPPFIKITYEELPLIRNKTLISGL 346
1C
1C RESULT 2
1C ID AAY41323 standard; protein; 347 AA.
1C AC AAY41323;
1C DT 02-DEC-1999 (first entry)
1C XX
1C DE Human secreted protein encoded by gene 16 clone HMZAD7.
1C XX
1C KM Human; secreted protein; fusion protein; gene therapy; protein therapy;
1C KM diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
1C KM developmental abnormality; foetal deficiency; blood; allergy; renal;
1C KM immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
1C KM inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
1C KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
1C KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
1C KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
1C OS
1C XX Homo sapiens.
1C PN MO9947540-A1.
1C PD 23-SEP-1999.
1C PF 18-MAR-1999; 99WO-US005804.
1C PR 19-MAR-1998; 98US-0078563P.
1C PR 19-MAR-1998; 98US-0078566P.
1C PR 19-MAR-1998; 98US-0078572P.
1C PR 19-MAR-1998; 98US-0078574P.
1C PR 19-MAR-1998; 98US-0078576P.
1C PR 19-MAR-1998; 98US-0078577P.

```

```

1C 19-MAR-1998; 98US-0078578P.
1C 19-MAR-1998; 98US-0078579P.
1C 19-MAR-1998; 98US-0078581P.
1C 01-APR-1998; 98US-0080312P.
1C 01-APR-1998; 98US-0080313P.
1C 01-APR-1998; 98US-0080314P.
1C
1C (HUMA-) HUMAN GENOME SCI INC.
1C
1C Ruben SM, Nij J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;
1C Wei Y, Endress GA, Duan RD, Kyaw H, Edner R, Lafleur DW, Olsen HS;
1C Shi Y, Moore PA;
1C WPI; 1999-562050/47.
1C N-PSDB; AA224826.
1C
1C New isolated human genes, useful for diagnosis and treatment of e.g.
1C cancers, neurological disorders, immune diseases, inflammation or blood
1C disorders.
1C
1C Claim 11; Page 367-368; 484pp; English.
1C
1C This sequence represents a secreted human protein encoded by the gene
1C clone detailed in the descriptor line. The gene can be used to generate
1C fusion proteins by linking to the gene to a human immunoglobulin Fc
1C portion (e.g. AA224802) for increasing the stability of the fused protein
1C as compared to the human protein only. The invention relates to 95 novel
1C genes and their fragments (nucleic acid sequences: AA224811-224907; amino
1C acid sequences AA41308-41404) which are useful for preventing, treating
1C or ameliorating medical conditions e.g. by protein or gene therapy. Also,
1C new polypeptides in a sample or by determining the amount of the
1C in the new polynucleotides. Specific uses are described for each of the
1C 95 polynucleotides, based on which tissues they are most highly expressed
1C in (see AA224811 for described uses)
1C
1C SQ
1C Sequence 347 AA:
1C
1C Query Match 99.4%; Score 1916; DB 2; Length 347;
1C Best Local Similarity 99.4%; Pred. No. 1.1e-184;
1C Matches 344; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1C
1C 1 MRGAGAGARGASGCMALALMLAVPGMSRVSGIPSRHMPVYKRPDRPKDPYCOA 60
1C 1 MRGAGAGARGASGCMALALMLAVPGMSRVSGIPSRHMPVYKRPDRPKDPYCOA 60
1C
1C 61 KYTFCPTGSPIPVWEGDDIEVFRLQAPVWFEFKYGDLLGHKIMHDAIGFRSTLTGKNT 120
1C 61 KYTFCPTGSPIPVWEGDDIEVFRLQAPVWFEFKYGDLLGHKIMHDAIGFRSTLTGKNT 120
1C
1C 121 MEWYELFQLGNCITFPHLRPEMDAPFMCNOGAACFEIGIDVHWKENGTLVQVATISGMF 180
1C 121 MEWYELFQLGNCITFPHLRPEMDAPFMCNOGAACFEIGIDVHWKENGTLVQVATISGMF 180
1C
1C 181 NQAKWKQKQNEGTIYETNNVKAPEKGAETWDSYDSCSFVLRRTNKALEFAERKNI 240
1C 181 NQAKWKQKQNEGTIYETNNVKAPEKGAETWDSYDSCSFVLRRTNKALEFAERKNI 240
1C
1C 241 ETNTTRIFLYSGEPTVGNESVGPCKNTGLAIRFYPPPHIPTKEFLLSLQIF 300
1C 241 ETNTTRIFLYSGEPTVGNESVGPCKNTGLAIRFYPPPHIPTKEFLLSLQIF 300
1C
1C 301 DAVIVHKQFYLFYNEFWFLPMKPPFIKITYEELPLIRNKTLISGL 346
1C 301 DAVIVHKQFYLFYNEFWFLPMKPPFIKITYEELPLIRNKTLISGL 346
1C
1C 301 DAVIVHKQFYLFYNEFWFLPMKPPFIKITYEELPLIRNKTLISGL 346
1C 301 DAVIVHKQFYLFYNEFWFLPMKPPFIKITYEELPLIRNKTLISGL 346
1C
1C RESULT 3
1C ID AAM93870 standard; protein; 358 AA.
1C AC AAM93870;
1C XX

```

```

06-NOV-2001 (first entry)
Human polypeptide, SEQ ID NO: 3978.
Human; full length cDNA; cDNA synthesis; oligo-capping.
Homo sapiens.
EP130094-A2.
05-SEP-2001.
07-JUL-2000; 2000EP-00114089.
08-JUL-1999; 99JP-00194486.
11-JAN-2000; 2000JP-00118774.
02-MAY-2000; 2000JP-00183765.
(HELI-) HELIX RES INST.
Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
WPI; 2001-524255/58.
N-PSDB; AAK94829.
830 Primers useful for synthesizing full length cDNA clones and their use
in genetic manipulation.
Claim 8; SEQ ID NO 3978; 1380bp + Sequence Listing; English.
The invention relates to primers for synthesizing full length cDNA
clones. 830 cDNA molecules encoding a human protein have been isolated
and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
been determined. Primers for synthesizing the full length cDNA are useful
for clarifying the function of the protein encoded by the cDNA. The full
length clones were obtained by construction of full length enriched cDNA
libraries that were synthesised by the oligo-capping method. The primers
enable the production of the full length cDNA easily without any special
methods. The present sequence is a polypeptide encoded by a full length
human cDNA of the invention. Note: The sequence data for this patent did
not form part of the printed specification, but was obtained in CD-ROM
format directly from EPO
Sequence 358 AA;
Query Match 99.3%; Score 1913; DB 4; Length 358;
Best Local Similarity 99.4%; Pred. No. 2.2e-184;
Matches 344; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 MRGGAAGRGRASGCMALALMLAVVPGMSKVSIGPSRAHMYVYKRPFRKPPYCOA 60
13 MRGGAAGRGRASGCMALALMLAVVPGMSKVSIGPSRAHMYVYKRPFRKPPYCOA 72
61 KYTFCGTGSPVMEGDDDDIEVFRLOAPYWEKYGDLGHLKIMDAIGFRSTLGKNT 120
73 KYTFCGTGSPVMEGDDDDIEVFRLOAPYWEKYGDLGHLKIMDAIGFRSTLGKNT 132
121 MEMYELFQNGCTPHLRREMDAPFMCNGAACFFEGIDDVHMKNGTLVQYATISGNMF 180
133 MEMYELFQNGCTPHLRREMDAPFMCNGAACFFEGIDDVHMKNGTLVQYATISGNMF 192
181 NOMAKWVNDNTGTYETWNVKASPEKGAETWDSYDCKSVLRTENKLAFFGAEFNKI 240
193 NOMAKWVNDNTGTYETWNVKASPEKGAETWDSYDCKSVLRTENKLAFFGAEFNKI 252
241 ENYNYRIFLYSEPTLYLGNETSVEGPTGNKTLGLAIKRFYYPFKPHLPKXEPILSLQIF 300
253 ENYNYRIFLYSEPTLYLGNETSVEGPTGNKTLGLAIKRFYYPFKPHLPKXEPILSLQIF 312
301 DAVIVKQFYLYFNFYWFLEPKFPFIKITTYEELPLPIRNKTLISGL 346
313 DAVIVKQFYLYFNFYWFLEPKFPFIKITTYEELPLPIRNKTLISGL 358

```

```

RESULT 4
ID AAM19604
AC AAM19604;
DT 12-OCT-2001 (first entry)
DE Peptide #6038 encoded by probe for measuring cervical gene expression.
KW Probe; human; microarray; gene expression; cervical epithelial cell;
cervical cancer.
OS Homo sapiens.
MO200157278-A2.
09-AUG-2001.
30-JAN-2001; 2001WO-US000670.
04-FEB-2000; 2000US-0180312P.
26-MAY-2000; 2000US-0207456P.
30-JUN-2000; 2000US-00608408.
03-AUG-2000; 2000US-00632386.
21-SEP-2000; 2000US-0234687P.
27-SEP-2000; 2000US-0236359P.
04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX DR
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX PS Claim 27; SEQ ID NO 24430; 487bp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX (SENP; see AAT10068-AAT28459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO atftp.wipo.int/pub/published_pct_sequences
SQ Sequence 163 AA;
Query Match 46.0%; Score 886; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 4.8e-81;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 AKWVQDNDETGYIETWNVKASPEKGAETWDSYDCKSVLRTENKLAFFGAEFNKI 243
61 YTRIFLYSGEPTLYLGNETSVEGPTGNKTLGLAIKRFYYPFKPHLPKXEPILSLQIF 303
244 YTRIFLYSGEPTLYLGNETSVEGPTGNKTLGLAIKRFYYPFKPHLPKXEPILSLQIF 346
304 YTRIFLYSGEPTLYLGNETSVEGPTGNKTLGLAIKRFYYPFKPHLPKXEPILSLQIF 358
121 YTRIFLYSGEPTLYLGNETSVEGPTGNKTLGLAIKRFYYPFKPHLPKXEPILSLQIF 163

```

RESULT 5

```
AB39239
D ABB39239 standard; peptide; 163 AA.
K ABB39239;
K
K 04-FEB-2002 (first entry)
K
K Peptide #6745 encoded by human foetal liver single exon probe.
K
K Human; foetal liver; gene expression; single exon nucleic acid probe.
K
K Homo sapiens.
K
K MO200157277-A2.
K
K 09-AUG-2001.
K
K 30-JAN-2001; 2001MO-US000669.
K
K 04-FEB-2000; 2000US-0180312P.
K 26-MAY-2000; 2000US-0207456P.
K 30-JUN-2000; 2000US-00608408.
K 03-AUG-2000; 2000US-00632366.
K 21-SEP-2000; 2000US-0234687P.
K 27-SEP-2000; 2000US-0236359P.
K 04-OCT-2000; 2000GB-00024263.
K
K (MOLE-) MOLECULAR DYNAMICS INC.
K
K Penn SG, Hanzel DK, Chen W, Rank DR;
K
K WPI; 2001-483447/52.
K
K Human genome-derived single exon nucleic acid probes useful for analyzing
K gene expression in human fetal liver.
K
K Claim 27; SEQ ID NO 31874; 639pp + Sequence listing; English.
K
K The invention relates to a single exon nucleic acid probe for measuring
K human gene expression in a sample derived from human foetal liver. The
K single exon nucleic acid probes may be used for predicting, measuring and
K displaying gene expression in samples derived from human fetal liver. The
K present sequence is a peptide encoded by a single exon nucleic acid probe
K part of the invention. Note: The sequence data for this patent did not form
K directly from WIPO at fip.wipo.int/pub/published_pcc_sequences
K
K Sequence 163 AA;
K
K Query Match 46.0%; Score 886; DB 4; Length 163;
K Best Local Similarity 100.0%; Pred. No. 4,8e-81;
K Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
K
K Y 184 AKWVKQDNETGIYETWNVKASPEKGAETWFDSDYCSKFLRTFNKLAERGAEPKNIETN 243
K 1 AKWVKQDNETGIYETWNVKASPEKGAETWFDSDYCSKFLRTFNKLAERGAEPKNIETN 60
K
K Y 244 YTRIFLYSGEPFYLGNETSFGPTGNKTLGLAIKRFYEPKPHLPKKEFLLSLLOIFDAY 303
K 61 YTRIFLYSGEPFYLGNETSFGPTGNKTLGLAIKRFYEPKPHLPKKEFLLSLLOIFDAY 120
K
K Y 304 IVHKQFYLFFNFEYWFLLPMKFPFIKITYEELPIPIRNKTLISGL 346
K 121 IVHKQFYLFFNFEYWFLLPMKFPFIKITYEELPIPIRNKTLISGL 163
K
K RESULT 6
K AAM32740 standard; protein; 163 AA.
K
K AAM32740;
K
K 17-OCT-2001 (first entry)
```

```
XX
K Peptide #6777 encoded by probe for measuring placental gene expression.
K
K Probe; microarray; human; placenta; antenatal diagnosis;
K genetic disorder.
K
K Homo sapiens.
K
K MO200157272-A2.
K
K 09-AUG-2001.
K
K 30-JAN-2001; 2001MO-US000663.
K
K 04-FEB-2000; 2000US-0180312P.
K 26-MAY-2000; 2000US-0207456P.
K 30-JUN-2000; 2000US-00608408.
K 03-AUG-2000; 2000US-00632366.
K 21-SEP-2000; 2000US-0234687P.
K 27-SEP-2000; 2000US-0236359P.
K 04-OCT-2000; 2000GB-00024263.
K
K (MOLE-) MOLECULAR DYNAMICS INC.
K
K Penn SG, Hanzel DK, Chen W, Rank DR;
K
K WPI; 2001-488937/53.
K
K Human genome-derived single exon nucleic acid probes useful for analyzing
K gene expression in human placenta.
K
K Claim 27; SEQ ID NO 33009; 654pp; English.
K
K The present invention relates to single exon nucleic acid probes (SENP:
K see A113135-A157546). The present sequence is a peptide encoded by one
K such probe. The probes are useful for producing a microarray for
K predicting, measuring and displaying gene expression in samples derived
K from human placenta. The probes are useful for antenatal diagnosis of
K human genetic disorders
K
K Sequence 163 AA;
K
K Query Match 46.0%; Score 886; DB 4; Length 163;
K Best Local Similarity 100.0%; Pred. No. 4,8e-81;
K Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
K
K QY 184 AKWVKQDNETGIYETWNVKASPEKGAETWFDSDYCSKFLRTFNKLAERGAEPKNIETN 243
K 1 AKWVKQDNETGIYETWNVKASPEKGAETWFDSDYCSKFLRTFNKLAERGAEPKNIETN 60
K
K QY 244 YTRIFLYSGEPFYLGNETSFGPTGNKTLGLAIKRFYEPKPHLPKKEFLLSLLOIFDAY 303
K 61 YTRIFLYSGEPFYLGNETSFGPTGNKTLGLAIKRFYEPKPHLPKKEFLLSLLOIFDAY 120
K
K QY 304 IVHKQFYLFFNFEYWFLLPMKFPFIKITYEELPIPIRNKTLISGL 346
K 121 IVHKQFYLFFNFEYWFLLPMKFPFIKITYEELPIPIRNKTLISGL 163
K
K RESULT 7
K ABB24080 standard; protein; 163 AA.
K
K ABB24080;
K
K 23-JAN-2002 (first entry)
K
K Protein #6079 encoded by probe for measuring heart cell gene expression.
K
K Human; gene expression; heart; microarray; vascular system;
K cardiovascular disease; hypertension; cardiac arrhythmia;
K congenital heart disease.
K
K
```


Homo sapiens.
 WO200157274-A2.
 09-AUG-2001.
 30-JAN-2001; 2001WO-US000666.
 04-FEB-2000; 2000US-0180312P.
 26-MAY-2000; 2000US-0207456P.
 30-JUN-2000; 2000US-00608408.
 03-AUG-2000; 2000US-00632366.
 21-SEP-2000; 2000US-0234687P.
 27-SEP-2000; 2000US-0236359P.
 04-OCT-2000; 2000GB-00024263.
 (MOLE-) MOLECULAR DYNAMICS INC.
 Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-48899/53.
 Single exon nucleic acid probes for analyzing gene expression in human hearts.
 Claim 15; SEQ ID NO 25850; 530bp; English.
 The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA21533-AM41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease. Note: This sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 163 AA;
 Query Match 46.0%; Score 886; DB 4; Length 163;
 Best Local Similarity 100.0%; Pred. No. 4.8e-81;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

184 AKWVKQDNETGTYETWVWKASPEKGAETWPSYDCKSVLTENKLAFFGAEFKNIETN 243
 1 AKWVKQDNETGTYETWVWKASPEKGAETWPSYDCKSVLTENKLAFFGAEFKNIETN 60
 244 YTRIFLYSGEPYLLGNETSVEFGPTGNKTGLAIKRFYYPFKDHLPTKEFLSLLOIFDAV 303
 61 YTRIFLYSGEPYLLGNETSVEFGPTGNKTGLAIKRFYYPFKDHLPTKEFLSLLOIFDAV 120
 304 IVHKQFYLFYNEFWFLPMKPFPIKITYEIPLPIRNKTLISGL 346
 121 IVHKQFYLFYNEFWFLPMKPFPIKITYEIPLPIRNKTLISGL 163

RESULT 8
 AAM72492
 AAM72492 standard; protein; 163 AA.
 AAM72492;
 06-NOV-2001 (first entry)
 Human bone marrow expressed probe encoded protein SEQ ID NO: 32798.
 Human; bone marrow expressed exon; gene expression analysis; probe;
 microarray; cancer; leukaemia; lymphoma; myeloma.
 Homo sapiens.

XX
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000668.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-48890/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human bone marrow.
 XX
 PS Example 4; SEQ ID NO 32798; 658bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention
 XX
 SQ Sequence 163 AA;
 Query Match 46.0%; Score 886; DB 4; Length 163;
 Best Local Similarity 100.0%; Pred. No. 4.8e-81;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

184 AKWVKQDNETGTYETWVWKASPEKGAETWPSYDCKSVLTENKLAFFGAEFKNIETN 243
 1 AKWVKQDNETGTYETWVWKASPEKGAETWPSYDCKSVLTENKLAFFGAEFKNIETN 60
 244 YTRIFLYSGEPYLLGNETSVEFGPTGNKTGLAIKRFYYPFKDHLPTKEFLSLLOIFDAV 303
 61 YTRIFLYSGEPYLLGNETSVEFGPTGNKTGLAIKRFYYPFKDHLPTKEFLSLLOIFDAV 120
 304 IVHKQFYLFYNEFWFLPMKPFPIKITYEIPLPIRNKTLISGL 346
 121 IVHKQFYLFYNEFWFLPMKPFPIKITYEIPLPIRNKTLISGL 163

RESULT 9
 AAM59901
 AAM59901 standard; protein; 163 AA.
 AAM59901;
 05-NOV-2001 (first entry)
 Human brain expressed single exon probe encoded protein SEQ ID NO: 32006.
 Human; brain expressed exon; gene expression analysis; probe; microarray;
 Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
 Homo sapiens.
 WO200157275-A2.
 09-AUG-2001.
 30-JAN-2001; 2001WO-US000667.
 PF
 XX

```

R 04-FEB-2000; 2000US-0180312P.
R 26-MAY-2000; 2000US-0207456P.
R 30-JUN-2000; 2000US-00608406.
R 03-AUG-2000; 2000US-00632366.
R 21-SEP-2000; 2000US-0234687P.
R 27-SEP-2000; 2000US-0236359P.
R 04-OCT-2000; 2000GB-00024263.
X
X (MOLE-) MOLECULAR DYNAMICS INC.
X Penn SG, Hanzel DK, Chen W, Rank DR;
X WPI; 2001-483446/52.
R
T Single exon nucleic acid probes for analyzing gene expression in human
T brains.
X
X Example 4; SEQ ID NO 32006; 650bp + Sequence Listing; English.
X
C The present invention provides a number of single exon nucleic acid
C probes which are derived from genomic sequences expressed in the human
C brain. They can be used to measure gene expression in brain cell samples,
C which may enable the diagnosis and improved treatment of nervous system
C diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
C epilepsy and cancers. The present sequence is a protein encoded by one of
C the probes of the invention
X
X Sequence 163 AA:
Q
Query Match 46.0%; Score 886; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 4.8e-81;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 184 AKWVKQDNERTGIYETWVNWKASPEKGAETWDSYDSKVLRTFNKLAERGAEFKNIEN 243
b 1 AKWVKQDNERTGIYETWVNWKASPEKGAETWDSYDSKVLRTFNKLAERGAEFKNIEN 60
Y 244 YTRIFLYSGEPTLYGNETSVEFGPTGNKTLGLAIKRFYFFPKPHLPYKEFLLSLQIFDAV 303
b 61 YTRIFLYSGEPTLYGNETSVEFGPTGNKTLGLAIKRFYFFPKPHLPYKEFLLSLQIFDAV 120
Y 304 IVHKQFYLFYNEFYWFLPMKFPPIKITTYEIRIPLPINKTSLG 346
b 121 IVHKQFYLFYNEFYWFLPMKFPPIKITTYEIRIPLPINKTSLG 163
X
X RESULT 10
X BG54183
X ABG54183 standard; peptide; 163 AA.
X
X ABG54183;
C
T 25-FEB-2003 (first entry)
X
X Human liver peptide, SEQ ID NO 32831.
X
X Human liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
X hypercholesterolaemia; coronary heart disease.
X
X Homo sapiens.
X
X WO200157273-A2.
X
X 09-AUG-2001.
X
X 30-JAN-2001; 2001WO-US000664.
X
X 04-FEB-2000; 2000US-0180312P.
X 26-MAY-2000; 2000US-0207456P.
X 30-JUN-2000; 2000US-00608406.
X 03-AUG-2000; 2000US-00632366.
X 21-SEP-2000; 2000US-0234687P.
X 27-SEP-2000; 2000US-0236359P.
X 04-OCT-2000; 2000US-0236359P.

```

```

PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488998/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
PT
XX Claim 27; SEQ ID NO 32831; 658bp; English.
PS
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 1109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABG47348-ABG59930 represent human
XX liver single exon encoded peptides of the invention. Note: The sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
X Sequence 163 AA:
SQ
Query Match 46.0%; Score 886; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 4.8e-81;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 184 AKWVKQDNERTGIYETWVNWKASPEKGAETWDSYDSKVLRTFNKLAERGAEFKNIEN 243
Db 1 AKWVKQDNERTGIYETWVNWKASPEKGAETWDSYDSKVLRTFNKLAERGAEFKNIEN 60
QY 244 YTRIFLYSGEPTLYGNETSVEFGPTGNKTLGLAIKRFYFFPKPHLPYKEFLLSLQIFDAV 303
Db 61 YTRIFLYSGEPTLYGNETSVEFGPTGNKTLGLAIKRFYFFPKPHLPYKEFLLSLQIFDAV 120
QY 304 IVHKQFYLFYNEFYWFLPMKFPPIKITTYEIRIPLPINKTSLG 346
Db 121 IVHKQFYLFYNEFYWFLPMKFPPIKITTYEIRIPLPINKTSLG 163
X
X RESULT 11
X ABG42312
X ID ABG42312 standard; peptide; 163 AA.
X
X ABG42312;
AC
X
X 19-AUG-2002 (first entry)
DT
XX
XX Human peptide encoded by genome-derived single exon probe SEQ ID 31977.
DE
XX
XX Human; single exon probe; asthma; lung cancer; COPD; IHD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemostasis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX
XX Homo sapiens.
XX
XX WO200186003-A2.
XX
XX 15-NOV-2001.
XX

```

30-JAN-2001; 2001WO-US000665.
 04-FEB-2000; 2000US-0180312P.
 26-MAY-2000; 2000US-0207456P.
 30-JUN-2000; 2000US-00608408.
 03-AUG-2000; 2000US-00632366.
 21-SEP-2000; 2000US-0234687P.
 27-SEP-2000; 2000US-0236359P.
 04-OCT-2000; 2000GB-00024263.
 (MOE-) MOLECULAR DYNAMICS INC.
 Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2002-114183/15.
 Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples.
 Claim 27; SEQ ID NO 31977; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample delivered from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsageners syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 163 AA;

Query Match 46.0%; Score 886; DB 5; Length 163;
 Best Local Similarity 100.0%; Pred. NO. 4.8e-81;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

184 AKWKQDNETGIYETWNVKASPEKGAETWDSYDCSFVIRTNKLAEFAEKNIETN 243
 1 AKWKQDNETGIYETWNVKASPEKGAETWDSYDCSFVIRTNKLAEFAEKNIETN 60
 244 YTRIFLVSGETPYIGNETSVFGPTGNKTLGAIKRFYYPKPHLPYKRFLLSLQIFDAV 303
 61 YTRIFLVSGETPYIGNETSVFGPTGNKTLGAIKRFYYPKPHLPYKRFLLSLQIFDAV 120

CY 304 IVHKQFYLFNFENYFWLPMKPFPIKITYEIPLPRNKLISGL 346
 DB 121 IVHKQFYLFNFENYFWLPMKPFPIKITYEIPLPRNKLISGL 163

RESULT 12
 ABP42902
 ID ABP42902 standard; protein, 89 AA.
 XX ABP42902;
 AC
 XX
 DT 22-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HPDQ081, SEQ ID NO:4034.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 XX ovarian cancer; breast cancer; tumour; reproductive system disorder;
 XX infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 XX PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
 XX inflammatory condition; immune disorder; blood disorder;
 XX cardiovascular disorder; respiratory disorder; neurological disorder;
 XX gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.

OS Homo sapiens.
 XX
 XX WO200200677-A1.
 XX
 XX 03-JAN-2002.
 XX
 XX 07-JUN-2001; 2001WO-US018569.
 XX
 XX 07-JUN-2000; 2000US-0209467P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Biese CE, Rosen CA;
 XX
 XX WPI; 2002-147878/19.
 DR N-PSDB; ABQ55979.
 XX

PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 PT cancer), immune disorders, cardiovascular disorders and neurological
 PT diseases.
 XX
 XX Claim 11; SEQ ID NO 4034; 2922pp; English.

XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovarian and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the

identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents a human ovarian antigen of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 89 AA;

Query Match 24.5%; Score 472; DB 5; Length 89;
Best Local Similarity 98.9%; Pred. No. 1,4e-39;
Matches 88; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

258 GNETSVGPNTKLTGLAIKRFYPPFKPHLPTEKFLISLQIPDAIVHKQFLTFYNFEX 317
1 GNETSVGPNTKLTGLAIKRFYPPFKPHLPTEKFLISLQIPDAIVHKQFLTFYNFEX 60

318 WFLPMKEPFIKITYEELIPLRNKTLISGL 346
61 WFLPMKEPFIKITYEELIPLRNKTLISGL 89

RESULT 13
ABR52348 standard; protein; 1705 AA.

ABR52348;
19-JUN-2003 (first entry)

Protein relating to the invention SEQ ID NO: 37.

antiproliferative; hepatotropic; nephrotropic; antiarthritic;
antiproliferative; cardiatic; cytostatic; gene therapy; liver disease;
proliferative disorder; renal failure; cardiovascular disorder;
immunological disorder; arthritis; psoriasis; congenital heart defect;
congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.

Mus musculus.
WO200257460-A2.
25-JUL-2002.

20-DEC-2001; 2001WO-US050459.
20-DEC-2000; 2000US-0256868P.
30-MAR-2001; 2001US-0280186P.
01-MAY-2001; 2001US-0287735P.
05-JUN-2001; 2001US-0295848P.
25-JUN-2001; 2001US-0300465P.

(BRIM) BRISTOL-MYERS SQUIBB CO.
Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L,
Siemens N, Bol D, Schieven G, Finger J, Toddard CG, Bassolino D,
Krystek S, Mcatee P, Suchard S, Barnes D;
WPI; 2002-599721/64.

Novel polynucleotides encoding human phosphatase polypeptides useful in the prevention or treatment of e.g. proliferative and cardiovascular disorders.

Disclosure, Fig 9; 80pp; English.

The invention relates to a novel isolated nucleic acid comprising a polynucleotide having a nucleotide sequence selected from 40 polynucleotides fully defined in the specification. The polynucleotide of the invention has antiproliferative, hepatotropic, nephrotropic, antiarthritic, antipsoriatic, cardiatic, and cytostatic activity. The polynucleotide may have a use in gene therapy. A polynucleotide or

polypeptide of the invention is useful for preventing, treating or ameliorating a medical condition, e.g. a proliferative disorder. They are also useful for treating e.g. liver disease, renal failure, immunological disorders including arthritis and psoriasis, cardiovascular disorders such as congenital heart defects and congestive heart failure, and cancer. A method of the invention is useful for diagnosing a pathological condition or susceptibility to a condition in a subject. The present sequence is used in the exemplification of the invention

Sequence 1705 AA;

Query Match 5.0%; Score 96.5; DB 5; Length 1705;
Best Local Similarity 21.4%; Pred. No. 9.8;
Matches 78; Conservative 40; Mismatches 134; Indels 113; Gaps 17;

2 RRGAG-----AARGASMCWALMLTAVVPGW-----SRVSGIPSR----- 39
32 RQGGPPPLSVNVSRRKPT--SLFLSWAABDGGDYALCLRAMNLGFPBGQJQAH 88

40 -----HMPVPYKRPD-----FRPKDPYCOAKYFC-----PTGSPT 71
89 NSSFFHGLVNGSRKYQLLVLR-----CWQNTVITLTARTAPVAVGLQHSIGSPA 143

72 PV-----MEGDDIEVFRLQAP-----VMEFKYGDILGHUKTMHDAIGFR 111
144 SLEASWSDASGDPSQYLLVHPESHTLACNVSVSPDTLSYNGDILPSQYVLEVTWA 203

112 STLTGKNYMEWELEQLNGCTEPFHLRPEMDAPFW--CNOGACFEFGIDVHKRENTLY 170
204 GSLHAFTSLQWTEPVPDPHILTRALGTSSLOAFMNSSRCATWFHLLTDLGSLTILKY 263

171 QVATISGNMFMQAKKWKODNETGIYY-----ETWVKASPEKGAETWFDSDYC 219
264 VRGISTHIFLRIS-----PGTPYQKICAAAGPHQWGNAT-----EWTYSYPS 310

220 SKVLTETFKLAEFGAEFKNIETNYRILYSG---EPTLYGNE--TSVVG---PTGNKT 271
311 DVLTLPLMELMWSWAGQARDGY--VTKLSGVENTTTLGPECNAAVPPGILPPGHYT 368

272 LGTAL 276
369 LGLRV 373

RESULT 14
ABB60074 standard; protein; 2703 AA.

ABB60074;
26-MAR-2002 (first entry)

Drosophila melanogaster polypeptide SEQ ID NO 7014.

Drosophila melanogaster polypeptide biology; cell signalling; insecticide; pharmaceutical.
Drosophila melanogaster.
WO200171042-A2.

27-SEP-2001.
23-MAR-2001; 2001WO-US009231.
23-MAR-2000; 2000US-0191637P.
11-JUL-2000; 2000US-00614150.

(PEKE) PE CORP NY.
Venter JC, Adams M, Li FWD, Myers EW;
WPI; 2001-656860/75.

1 N-PSDB; AB104177.
2
3 New isolated nucleic acid detection reagent for detecting 1000 or more
4 genes from Drosophila and for elucidating cell signaling and cell-cell
5 interactions.
6
7 Disclosure; SEQ ID NO 7014; 21bp + Sequence Listing; English.
8
9 The invention relates to an isolated nucleic acid detection reagent
10 capable of detecting 1000 or more genes from Drosophila. The invention is
11 useful in developmental biology and in elucidating cell signaling and
12 cell-cell interactions in higher eukaryotes for the development of
13 insecticides, therapeutics and pharmaceutical drugs. The invention
14 discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
15 sequences (AB101840-AB116175) and the encoded proteins (AB057737-
16 AB072072). The sequence data for this patent did not form part of the
17 printed specification, but was obtained in electronic format directly
18 from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://pub.int/pub/published_pct_sequences)
19
20 Sequence 2703 AA;
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

4 protein - protein search, using sw model

on: June 8, 2004, 14:00:33 ; Search time 26.9111 Seconds
(without alignments)
663.762 Million cell updates/sec

title: US-10-010-050a-2
affect score: 1927
sequence: 1 MRRGAGARGRASWCWALAL.....IKITYEELPIPRNKTLISGL 346

scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 389414 segs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

set-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase :
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	91	4.7	4563	US-09-108-006C-1	Sequence 1, Appli
2	88.5	4.6	1462	US-07-792-600-31	Sequence 31, Appl
3	88.5	4.6	1462	US-09-157-021-31	Sequence 31, Appl
4	88.5	4.6	1462	US-09-156-842-31	Sequence 31, Appl
5	88.5	4.6	1462	US-09-591-514-31	Sequence 31, Appl
6	88	4.6	608	US-09-489-039A-13503	Sequence 13503, A
7	87.5	4.5	178	US-09-134-000C-3624	Sequence 3624, Ap
8	87.5	4.5	876	US-09-254-352B-19	Sequence 19, Appl
9	83	4.3	555	US-09-134-078-24	Sequence 24, Appl
10	83	4.3	1056	US-09-079-030-217	Sequence 217, Appl
11	83	4.3	1310	US-08-989-299-10	Sequence 10, Appl
12	83	4.3	1310	US-09-407-422B-27	Sequence 27, Appl
13	83	4.3	4536	US-09-180-422B-27	Sequence 27, Appl
14	83	4.3	4536	US-09-079-030-1	Sequence 1, Appli
15	82	4.3	322	US-09-252-991A-26565	Sequence 26565, A
16	81	4.2	741	US-09-489-039A-7854	Sequence 7854, Ap
17	81	4.2	771	US-09-462-284-2	Sequence 2, Appli
18	80.5	4.2	718	US-09-328-352-4640	Sequence 4640, Ap
19	80	4.2	451	US-09-489-039A-11849	Sequence 11849, A
20	80	4.2	607	US-09-537-682-1	Sequence 1, Appli
21	80	4.2	652	US-09-489-039A-12461	Sequence 12461, A
22	79.5	4.1	197	US-09-489-039A-8317	Sequence 8317, Ap
23	79	4.1	844	US-09-564-805-227	Sequence 227, Appl
24	78.5	4.1	634	US-09-849-334-4	Sequence 4, Appli
25	78.5	4.1	634	US-10-274-878-4	Sequence 4, Appli
26	78.5	4.1	669	US-09-134-000C-6185	Sequence 6185, Ap
27	77.5	4.0	626	US-08-956-242-2	Sequence 2, Appli

28	77.5	4.0	626	US-09-351-215-2	Sequence 2, Appli
29	77	4.0	198	US-09-252-991A-19295	Sequence 19295, A
30	77	4.0	238	US-08-634-475-6	Sequence 6, Appli
31	77	4.0	238	US-09-709-791-6	Sequence 6, Appli
32	77	4.0	405	US-09-134-001C-5259	Sequence 5259, Ap
33	77	4.0	501	US-08-781-802-4	Sequence 4, Appli
34	77	4.0	501	US-08-781-802-6	Sequence 6, Appli
35	77	4.0	501	US-08-694-078-4	Sequence 4, Appli
36	77	4.0	501	US-08-694-078-6	Sequence 6, Appli
37	77	4.0	501	US-09-058-260-4	Sequence 4, Appli
38	77	4.0	501	US-09-058-260-6	Sequence 6, Appli
39	77	4.0	501	US-09-058-260-14	Sequence 14, Appli
40	77	4.0	501	US-09-058-260-20	Sequence 20, Appli
41	77	4.0	501	US-09-058-260-22	Sequence 22, Appli
42	77	4.0	501	US-09-058-260-24	Sequence 24, Appli
43	77	4.0	501	US-09-058-260-32	Sequence 32, Appli
44	77	4.0	615	US-09-107-532A-6507	Sequence 6507, Ap
45	77	4.0	619	US-09-107-532A-4897	Sequence 4897, Ap

ALIGNMENTS

RESULT 1
US-09-108-006C-1
; Sequence 1, Application US/09108006C
; Patent No. 6524613
GENERAL INFORMATION:
APPLICANT: Steer, Clifford J.
Kren, Betsy T.
Bandyopadhyay, Paramita
Roy-Chowdhury, Jayanta
TITLE OF INVENTION: Hepatocellular Chimera-plasty
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSER: Kimeragen, Inc.
STREET: 300 Pheasant Run
CITY: Newtown
STATE: PA
COUNTRY: USA
ZIP: 18940
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09108, 006C
FILING DATE: 30-Jun-1992
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/054,288
FILING DATE: 30-APR-1997
APPLICATION NUMBER: 60/054,837
FILING DATE: 05-AUG-1997
APPLICATION NUMBER: 60/064,996
FILING DATE: 10-NOV-1997
APPLICATION NUMBER: 60/074,497
FILING DATE: 12-FEB-1998
APPLICATION NUMBER: PCT US 98/08834
FILING DATE: 30-APR-1998
ATTORNEY/AGENT INFORMATION:
NAME: Friebe, Thomas
REGISTRATION NUMBER: 29258
REFERENCE/DOCKET NUMBER: 7991-015-999
TELEPHONE: 215-504-4545
TELEFAX: 215-504-4444
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4563 amino acids
TYPE: amino acid

[illegible]

```

; LENGTH: 1462
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-591-514-31

Query Match          4.6%; Score 88.5; DB 4; Length 1462;
Best Local Similarity 20.3%; Pred. No. 2.5;
Matches 71; Conservative 56; Mismatches 115; Indels 107; Gaps 20;

QY  KRFPDPKPDYCOAKYTCPTG-----SPIPWEGD--DIEV---FRL 85
Db  KKRSIGASPNFVSHTTAVPSGKIASFVSRKEPPLTFVPLKRAFPADDDQVESTEEBQ 242
QY  QAPWEEFYGGLLHKLIMDAIGFRSTLTGKNTMYELFOLGNCFFPHRPEMDAF 145
Db  ESGAMEFEDGDF-----DERMEVEV-----DLEP-MAAKA 272
QY  MCNGAACFFEGCIDVHNKEN---GLIVQVATISGNMNMQAKM-VKODNENGIYYETMN 201
Db  WDKS-----EPAAEVKQKADSGKTV---SYLSEFLPDVSCWDIDEGSSFSVQEVQ 323
QY  VKASP--EKGAEI---WPSYD---CSKFLATFNKLAEGAE-----FNKI 240
Db  VDSHLLPLVKADEQVHFYMDAYEDQYNQPGVFLFGKWIASATHTVSCCYMVNVI 383
QY  ETNTRIFLYSGEPTLYGNETSVEFGTGNKT-LGIAIRFYYPEKPHLPTFEELSLIOI 299
Db  ERT-----LY-----FLPREKKIDLNTKETGTPTISMKDVVEEFDEKATATYKIMK--- 429
QY  FDVAIVHKGFLFYNEFYWFLPMKPEPFIKY-----EIPLEIRNKTLIS 344
Db  FKSKPVEKN---YAFELPDVPEKSEYILEVAKISAMFQLPDOLKGETPS 474

RESULT 6
US-09-489-039A-13503
; Sequence 13503, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13503
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13503

Query Match          4.6%; Score 88; DB 4; Length 608;
Best Local Similarity 20.9%; Pred. No. 0.7;
Matches 67; Conservative 41; Mismatches 95; Indels 118; Gaps 21;

QY  DIEFRLQAPWEEKYGGDLHKLIMDA--IGFRSTLTGK-NTYME---NYELFOLGNC 132
Db  DTEYRRVDP---QFGDAALLRLRHNTORAGMMIIDGVFNHTGDSHPWDRHQSSG 270
QY  TFPPIRPEMDAF-----WCNGAAC-----FEEGIDV--HWK 164
Db  KAPMSMDGWRILDVVMHLEGGAARNNLQIHIGITQAAKQAPFAFVCEHGDARQWLQA 387
QY  165 E-----NGTLVQVATIS-----GNMRYQAKYWKQ 189
Db  KAPMSMDGWRILDVVMHLEGGAARNNLQIHIGITQAAKQAPFAFVCEHGDARQWLQA 387
QY  190 D-----NETGIYETW-----NVKASPEK-GAET---WPSYDSCS---KFLVLTFNKL 230
Db  DAEDAAWNYRFTFPINGELANTIDISTDPQKIDQOTCMAMMDNTRAGISHQOOLRMENOL 447

```


Y 231 -----ABF-----GAERKNIETNYTRIFLYSGEP-TYLGNETSVFG---PTGNKTLGLAIK 277
 b 448 DSHDTARFKSLGKDVARDLAVVWLFPMGVGVCITYGGEVGVGDNNDPCKRP----- 501
 Y 278 RFFYPKPKHLPTEKEFLISLIQ 298
 b 502 ---FPWDPALQDYO-LLALYQ 518

RESULT 7

S-09-134-000C-3624
 Sequence 3624, Application US/09134000C
 Patent No. 6617156
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stramm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 032796-032
 CURRENT APPLICATION NUMBER: US/09/134,000C
 PRIOR FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/055,778
 PRIOR FILING DATE: 1997-08-15
 NUMBER OF SEQ ID NOS: 6812
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 3624
 LENGTH: 178
 TYPE: PR1
 ORGANISM: Enterococcus faecalis
 S-09-134-000C-3624

Query Match 4.5%; Score 87.5; DB 4; Length 178;
 Best Local Similarity 21.6%; Pred. No. 0.12; Matches 41; Conservative 24; Mismatches 60; Indels 65; Gaps 3;
 Y 117 KNYTMEWELFQJGNCPTPLRPMDAPFCWCGAACFEGCIDDVHKENGITVOYATIS 175
 b 13 KNYTRLRGERTDLSKGYISQLEKRLSP-----SWETFF-----SILEVLGYT 55
 Y 177 GNNENGM-----AKWQKDNETGIYVE-----TNVVKASPEKGATWFDSDYDCSFYAR 225
 b 56 PPOFESQQLPEQKIVNEBSTIYDEHGYELKMLIPASNEKEMP-----VII 103
 Y 226 TENKLAEPGAEPKNIETNYTRIFLY-----SGEPTYLGNETSVFGPTGNKTLGLAI 275
 b 106 TPDK-----NGEYKTEPSPJSETPFIYIDISIRLTGAEATYIAKK-----GQT----- 143
 Y 277 KREYYPFKPH 286
 b 149 -MYQATEPH 157

RESULT 8

S-09-254-352B-19
 Sequence 19, Application US/09254352B
 Patent No. 6365350
 GENERAL INFORMATION:
 APPLICANT: HAYASHIZAKI, Yoshinide
 TITLE OF INVENTION: METHOD OF DNA SEQUENCING
 FILE REFERENCE: 024705-080
 CURRENT APPLICATION NUMBER: US/09/254,352B
 PRIOR FILING DATE: 1999-11-10
 PRIOR APPLICATION NUMBER: PCT/JP98/03039
 PRIOR FILING DATE: 1998-07-06
 PRIOR APPLICATION NUMBER: JP 10-155847
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: JP 9-196478
 PRIOR FILING DATE: 1997-07-07
 NUMBER OF SEQ ID NOS: 64
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 19
 LENGTH: 876
 TYPE: PR1

ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: T7 RNA polymerase
 US-09-254-352B-19

Query Match 4.5%; Score 87.5; DB 4; Length 876;
 Best Local Similarity 21.6%; Pred. No. 1.4;
 Matches 63; Conservative 28; Mismatches 107; Indels 93; Gaps 17;

QY 1 MRGGAARGRASW---CWALLMLAVPG--WS-----RVSGIPSR---IHVPVY 45
 Db 267 LSKRGAALAGISPMHQPCVPPKPVWETVGGYWSGVALVTHSKALKRVDVHMPVY 326
 QY 46 KRDFRPPKDPYCOAK-----YTPCPGSPDPVMEGD-----DDIEVFRQAQV 89
 Db 327 KAVNLAQNTPMKVKKVLAVNVEIYNNMGCPYGD--VPALERELPRPPDIDTNEVARKA 385
 QY 90 WEKFGDILGHUKM-----HDAIGRSTLTGKNYMEW---YELP 127
 Db 386 WEKEAAAVYRKDKARQSRRCFCEFWVAQNFANHKAIME-----PYNMDWRGRVYAVS 438
 QY 128 QJGN-----CTPHLRP-EMDAPFWCN-OGACFEGSIDDVHKENGITVOYATISGN 178
 Db 439 MFGNDMTKGSILITLAKKPTIGLDGFYWLKIHGANC--AGVDVPPPE---RIKITEEN 491
 QY 179 MFWOMAK-----WVKQDNETGIYETMNVKASPEKGATWFDSDYDCS 220
 Db 492 EGNILASADPLNNTWWTQODSPFAFCFPAVAK-----HNGINSYNCS 535

RESULT 9

US-09-134-078-24
 Sequence 24, Application US/09134078
 Patent No. 6368844
 GENERAL INFORMATION:
 APPLICANT: Bylina, Edward J.
 TITLE OF INVENTION: GLYCOSIDASE ENZYMES
 NUMBER OF SEQUENCES: 72
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Gray Cary Ware & Freidenrich LLP
 STREET: 4365 Executive Drive, Suite 1600
 CITY: San Diego
 STATE: CA
 COUNTRY: USA
 ZIP: 92121
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows95
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/134,078
 FILING DATE: 13-AUG-1998
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/949,026
 FILING DATE: 10-OCT-1997
 APPLICATION NUMBER: 60/056,916
 FILING DATE: 06-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Haille, Lisa A.
 REGISTRATION NUMBER: 38,347
 REFERENCE/DOCKET NUMBER: 09010/024002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 858/677-1456
 TELEFAX: 858/677-1465
 INFORMATION FOR SEQ ID NO: 24:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 555 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal

b 276 INRGPIPAHLGNMMAQSWESIYDM-VPEPDP-KPNLDV----- 313
Y 164 KENGTLYOVATTISGNMFWQAKWYKQDNETGIYETW--NVKASPEKG-----AETWPD 215
b 314 --TSTWQKGMNATHMRVAAEEFTSLGLLPMPEFAESMLEKPEDGRVAVCHASAM-D 370
Y 216 SYDSCSKVLTFFNKLAFFGAEPKNIETNYRIPL-VSGEPPTYLGNETSVEPGTKNTLGL 274
b 371 FYNRKDFRIKQCTQVMTDQSLTVHHEMGHVQYLLQYKQDQVSLRRANPGFHEAIGVIAL 430
Y 275 AIKRFYFFPFIH-----PTKEFLSLLOI-----FDVAVIHKQFYL----- 312
b 431 SVSTPAHAKTIGLDHYNTDESINYLKMALEKIAFLPFGYLVQWQWGVSGRTPSG 490
Y 313 -YNEFEWFLPMKFPFIKITYEELPLIRNKT 342
b 491 RYNEFWMLRTKYQGI-----CPVVRNET 515

ESTULT 12

S-09-407-427-10

Sequence 10, Application US/09407427

Patent No. 6610497

GENERAL INFORMATION:

APPLICANT: Acton, Susan L.

APPLICANT: Robison, Keith E.

TITLE OF INVENTION: ANGIOGENSIN CONVERTING ENZYME HOMOLOG AND THERAPEUTIC

FILE REFERENCE: MNT-132CP2

CURRENT APPLICATION NUMBER: US/09/407,427

CURRENT FILING DATE: 1999-09-29

PRIOR APPLICATION NUMBER: 09/163,648

PRIOR FILING DATE: 1998-09-30

PRIOR APPLICATION NUMBER: 08/989,299

PRIOR FILING DATE: 1997-12-11

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 10

LENGTH: 1310

TYPE: PR1

ORGANISM: Oryctolagus cuniculus

S-09-407-427-10

Query Match 4.3%; Score 83; DB 4; Length 1310;

Best Local Similarity 19.7%; Pred. No. 8.8;

Matches 77; Conservative 61; Mismatches 145; Indels 108; Gaps 23;

Y 15 CML-----ALLMLAVVGSVSGISRRHVPVYKRFPRPDPYQQA 60
b 170 CWSLDPDLNLTASSRSYAML-LFAMEGMINAVGILPK--PL-YQEFALSN-EAYRQD 223
Y 61 KYTFCFPGS-----PIPVNEGDDIE-VERLQAPVWEFKYGDLLGLH-KIMMDALGFR- 111
b 224 --GSDTGAARWSMYDSPTFE--EDLERIYHQLEPL--YINLHVVRVILHRRYGDGR 275
Y 112 -----STLTGKNYTMWYELFOLGNCCTFPHLRPEMDAPFWQNOGAACFFEGIDVHM 163
b 276 INRGPIPAHLGNMMAQSWESIYDM-VPEPDP-KPNLDV----- 313
Y 164 KENGTLYOVATTISGNMFWQAKWYKQDNETGIYETW--NVKASPEKG-----AETWPD 215
b 314 --TSTWQKGMNATHMRVAAEEFTSLGLLPMPEFAESMLEKPEDGRVAVCHASAM-D 370
Y 216 SYDSCSKVLTFFNKLAFFGAEPKNIETNYRIPL-VSGEPPTYLGNETSVEPGTKNTLGL 274
b 371 FYNRKDFRIKQCTQVMTDQSLTVHHEMGHVQYLLQYKQDQVSLRRANPGFHEAIGVIAL 430
Y 275 AIKRFYFFPFIH-----PTKEFLSLLOI-----FDVAVIHKQFYL----- 312
b 431 SVSTPAHAKTIGLDHYNTDESINYLKMALEKIAFLPFGYLVQWQWGVSGRTPSG 490
Y 313 -YNEFEWFLPMKFPFIKITYEELPLIRNKT 342

Db 491 RYNEFWMLRTKYQGI-----CPVVRNET 515

RESULT 13

US-09-180-422B-27

Sequence 27, Application US/09180422B

Patent No. 6444644

GENERAL INFORMATION:

APPLICANT: BRUCKDORFER, KARL R

ETTELBALE, CAMILIE

TITLE OF INVENTION: ANTICOAGULANT PEPTIDE FRAGMENTS DERIVED

FROM APOLOPROTEIN B-100

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHIE, P.C.

STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/180,422B

FILING DATE: 07-Dec-1998

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B.J.

REGISTRATION NUMBER: 36663

REFERENCE/DOCKET NUMBER: 117-268

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7038164000

TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 4536 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 27:

US-09-180-422B-27

Query Match 4.3%; Score 83; DB 4; Length 4536;

Best Local Similarity 19.7%; Pred. No. 63;

Matches 56; Conservative 39; Mismatches 123; Indels 66; Gaps 11;

Y 74 MEGDDDIIEVFRLOAPVWEFKYGDLLG--HLKIMMDALGFRSLTGKNTMWEYELFQGL 130
b 3986 KQEDDDFS-----KKNFYSPQSSPDKKLTFTKELVRESDEDTQIKVNEEELAAAG 4038
Y 131 NCTFPHLRPEMDAPFWQNOGAACFFEGIDVHKENG-TLVQVATISGNMFWQAKWYKQ 189
b 4039 LFL-SIKQNV-----KATGVLDYVWKYWEHNTGLTBREVSLSLRNLQNNMAEVMYQ 4090
Y 190 DNETGIYETWNVKASPEKGAEFTWSDYDSCSFVLTENKLAFFGAEPKNIETN-YTRI 248
b 4091 ---GAIRQIDIDIVAFQKAAAGTGTGYO-----EMDKAQNLTQLEL 4129
Y 249 LVSGETYLGNETSVGPTGNKTLGLAIF-----REYVFFKPHLPTKEFL 293
b 4130 TOEGQASFOGLKDNVDFGLVTRTQKHHMKVLAISLDLFLNPPRQFQFGKPGIYRREL 4189
Y 294 LSLQIFDAVIVHKQFY-----LFYNEFEWFLPMKFPFIK 328
b 4190 CMMF-IREVGTVLSQYYSKVHNGSEILFSFYQDLVITLPEFLRK 4232

RESULT 14

US-09-079-030-1

Sequence 1, Application US/09079030
Patent No. 6635623

GENERAL INFORMATION:

APPLICANT: Guevera, Jr., Juan G.
APPLICANT: Hoogeveen, Ron C.

APPLICANT: Moore, Paul J.

TITLE OF INVENTION: LIPOPROTEINS AS NUCLEIC ACID DELIVERY
VECTORS FOR TRANSFECTION OF EUKARYOTIC CELLS

NUMBER OF SEQUENCES: 229

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/079,030

FILING DATE: Concurrently Herewith

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: McMillian, Nabeela R.

REGISTRATION NUMBER: P-43,363

REFERENCE/DOCKET NUMBER: ARAG:003

TELECOMMUNICATION INFORMATION:

TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4536 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

-09-079-030-1

Query Match 4.3%, Score 83; DB 4; Length 4536;

Best Local Similarity 19.7%; Pred. No. 63;

Matches 56; Conservative 39; Mismatches 123; Indels 66; Gaps 11;

```

74 MEGDDIEVFLQAPVPEFKYGDLLG--HLKIMHDAIGFRSTLTGKNYTWMEYELFOLG 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3986 MDEDDDS-----KMFYSPQSSPDCKLTITKTELRYVESDETOIKVWEEBAAG 4038
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
131 NCTPPLRPEKADAPFCNCGAACFFEGIDVHWKENG-TLVQVATTISGNMFMQAKWKQ 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4039 LLT--SLKDNV-----KATGVLYDYVKNKYHWEHTGLTLREVSSKLRNLONNAAEWYQ 4090
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 DNEGTLYETWNVKASPEKGAETWFDSDYDCKFVLRTPNKLAEFGAEKNTETN-YTRIF 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4091 ---GAIRQIDIDIVRFQKASGTTGTQY-----FWKDKACMLYQELL 4129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249 LYSGEPTLGNETSVPFGTKTKTGLAIK-----RFYPPKPHLPTEKEL 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4130 TQEGQASQGLKDNVDFGLVAVTQKFNKVKHLIDSLIDFNPRFQPGKPGIYTRBEL 4189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
294 LSLQIFDAVIVHKQFY-----LFYNEFWELPMKKEPFK 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4190 CTMP-IRVGVTLQGVYSKVNGSEILFSYFQDLVITLPPFLRK 4232
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

SUIT 15

-09-252-991A-26565

Sequence 26565, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26565
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26565

```

Query Match 4.3%, Score 82; DB 4; Length 322;

Best Local Similarity 25.0%; Pred. No. 1.2;

Matches 41; Conservative 8; Mismatches 59; Indels 56; Gaps 6;

```

QY 24 AVVPGMRVSGIPSRRLHPVYKRFDFRKPDPYCGAKYTCPTGSP1PVMEGDDIEVF 83
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 186 ALVGLREVGVVFGDMWGNLYPRSGFLHQTDY-----KTAIVIAQRAGDITTVVG 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 84 RLQA-----FWEFKYGDLLGLKIMHDAIGFRSTLTGKNYTWMEYEL---F 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 QLVHYLPMAAPKQGVWPAGELKEG-----ASTGK-----WQELTPSL 276
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 128 QLNGCTFPHLRPEMDA-----PFCNCGAACFFEGIDD 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 277 SLNCAVFPNSGPKTQAVDGDYAMALMRPYSCCQKKGQIFLGSTD 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: June 8, 2004, 14:09:15
Job time : 27.911 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - protein search, using sw model

on: June 8, 2004, 14:04:04 ; Search time 71.1222 Seconds
(without alignments)
1368.675 Million cell updates/sec

file: US-10-010-050a-2

Sequence: 1 MRGAGARGASWCMALAL.....IKTYEIPLPTRKTTSGL 346

oring table: BLOSUM62

Gapped 10.0 , Gapped 0.5

Number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCF_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/CTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	1927	100.0	346	9	US-09-122-383-2
2	1927	100.0	346	13	US-10-010-050a-2
3	1916	99.4	346	12	US-10-653-595-126
4	1916	99.4	347	12	US-09-397-945-126
5	886	46.0	163	9	US-09-864-761-39378
6	472	24.5	89	15	US-10-264-049-4034
7	97.5	5.1	328	15	US-10-369-493-12714
8	96	5.0	290	15	US-10-369-493-3172
9	95.5	5.0	836	16	US-10-408-765a-58
10	94	4.9	1238	12	US-10-282-122a-56749
11	93	4.8	478	12	US-10-282-122a-56615
12	91.5	4.7	347	12	US-10-424-599-262855
13	91.5	4.7	350	12	US-10-425-114-57996
14	91	4.7	742	10	US-09-374-046a-160
15	91	4.7	742	12	US-10-616-263-160

16	91	4.7	742	16	US-10-408-765a-2332	Sequence 2332, App
17	91	4.7	4563	9	US-09-870-758-128	Sequence 128, App
18	91	4.7	4563	10	US-09-751-708a-128	Sequence 128, App
19	89	4.6	495	12	US-10-424-599-252265	Sequence 252265, App
20	89	4.6	567	14	US-10-133-797-4	Sequence 4, Appl1
21	88	4.6	419	14	US-10-028-245-2	Sequence 2, Appl1
22	88	4.6	605	12	US-10-282-122a-60098	Sequence 60098, A
23	88	4.6	3194	12	US-10-282-122a-59046	Sequence 59046, A
24	87	4.5	1160	15	US-10-369-493-12860	Sequence 12860, A
25	86	4.5	411	12	US-10-282-122a-60577	Sequence 60577, A
26	85.5	4.4	483	12	US-10-425-114-43218	Sequence 43218, A
27	85.5	4.4	483	12	US-10-425-114-63234	Sequence 63234, A
28	84.5	4.4	224	16	US-10-461-990-35	Sequence 35, Appl
29	84.5	4.4	513	12	US-10-425-114-46548	Sequence 46548, A
30	84.5	4.4	513	12	US-10-425-114-66071	Sequence 66071, A
31	84	4.4	346	12	US-10-425-114-45828	Sequence 45828, A
32	84	4.4	356	12	US-10-424-599-225305	Sequence 225305, A
33	84	4.4	719	12	US-10-425-114-56258	Sequence 56258, A
34	83.5	4.3	373	12	US-10-012-819-198	Sequence 198, App
35	83.5	4.3	434	15	US-10-084-749-2336	Sequence 2336, App
36	83.5	4.3	984	15	US-10-369-493-8357	Sequence 8357, App
37	83	4.3	552	9	US-09-909-464a-2	Sequence 2, Appl1
38	83	4.3	555	13	US-10-121-032-24	Sequence 24, Appl
39	83	4.3	555	14	US-10-093-037-24	Sequence 24, Appl
40	83	4.3	1086	15	US-10-369-493-5963	Sequence 5963, App
41	83	4.3	4563	10	US-09-802-640-32	Sequence 32, Appl
42	83	4.3	4563	12	US-10-403-902a-32	Sequence 32, Appl
43	82.5	4.3	381	14	US-10-032-585-7666	Sequence 7666, App
44	82.5	4.3	591	14	US-10-081-872-156	Sequence 156, App
45	82.5	4.3	591	15	US-10-385-305-156	Sequence 156, App

ALIGNMENTS

```
RESULT 1
US-09-122-383-2
; Sequence 2, Application US/09122383a
; Patent No. US20020042093A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
; TITLE OF INVENTION: CHROMOSOME 13
; FILE REFERENCE: 97-38
; CURRENT APPLICATION NUMBER: US/09/122,383A
; EARLIER FILING DATE: 1998-07-24
; EARLIER APPLICATION NUMBER: 60/053,613
; EARLIER FILING DATE: 1997-07-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-122-383-2
```

Query Match	100.0%	Score 1927	DB 9	Length 346
Best Local Similarity	100.0%	Pred. No. 6e-181		
Matches 346	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MRGAGARGASWCMALALIMLAVVGVGSRVSGIPSRHWPVYKRFDFPKDPYCOA	60	
DB	1	MRGAGARGASWCMALALIMLAVVGVGSRVSGIPSRHWPVYKRFDFPKDPYCOA	60	
QY	61	KTYFCPTGSPDPVWEGDDIEVFRLQAPVWEFKYGDILGHUKIMHDAIGFSTLTGKYNT	120	
DB	61	KTYFCPTGSPDPVWEGDDIEVFRLQAPVWEFKYGDILGHUKIMHDAIGFSTLTGKYNT	120	
QY	121	MEWEFLFOLNCTFPHLRPEMDAPFVNCQACFFEGIDVHWKNGTIVQVATISGMF	180	
DB	121	MEWEFLFOLNCTFPHLRPEMDAPFVNCQACFFEGIDVHWKNGTIVQVATISGMF	180	

181 NQAKWVQDNDEGTGYETWNVKASPEKGATWTDSDSCFVLRITNKLAEGFAEFKNI 240
 181 NQAKWVQDNDEGTGYETWNVKASPEKGATWTDSDSCFVLRITNKLAEGFAEFKNI 240
 241 ETTNTRIFLYSGEPYLGNETSVFGPTGNKTLGLAIKRFYFPKPHLPTEFLLSLQIF 300
 241 ETTNTRIFLYSGEPYLGNETSVFGPTGNKTLGLAIKRFYFPKPHLPTEFLLSLQIF 300
 301 DAVIVHKQFYLFYNFEYWFLLPMKFPFIKITYEELPLPINKTSLGL 346
 301 DAVIVHKQFYLFYNFEYWFLLPMKFPFIKITYEELPLPINKTSLGL 346

RESULT 2
 S-10-010-050A-2

Sequence 2, Application US/10010050A
 Publication No. US20020173624A1

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

APPLICANT: Gilbertson, Debra G.

TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN

TITLE OF INVENTION: CHROMOSOME 13

FILE REFERENCE: 97-38C1

CURRENT APPLICATION NUMBER: US/10/010,050A

CURRENT FILING DATE: 2002-03-26

PRIOR APPLICATION NUMBER: US 09/122,383

PRIOR FILING DATE: 1998-07-24

PRIOR APPLICATION NUMBER: US 60/053,613

PRIOR FILING DATE: 1997-07-24

NUMBER OF SEQ ID NOS: 19

SOFTWARE: fastseq for windows version 4.0

SEQ ID NO 2

LENGTH: 346

TYPE: PRT

ORGANISM: Homo sapien

S-10-010-050A-2

Query Match 100.0%; Score 1927; DB 13; Length 346;
 Best Local Similarity 100.0%; Pred. No. 66-181;

Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRGAGAAARASWCMALALMLAVVPGSRVSGIPSRHWPVYKRFPRPDPYCOA 60
 1 MRGAGAAARASWCMALALMLAVVPGSRVSGIPSRHWPVYKRFPRPDPYCOA 60
 61 KYTFCPTGSPIPVMBEGDDIEVFRLQAPVWEFRYGDLLGLKIMHDAIGFRSTLTGKNT 120
 61 KYTFCPTGSPIPVMBEGDDIEVFRLQAPVWEFRYGDLLGLKIMHDAIGFRSTLTGKNT 120
 121 MEWYELFQJGNCCTPHLRPEMDAPFWCNGAACFFEGIDVHWKENGTLVQVATISGNMF 180
 121 MEWYELFQJGNCCTPHLRPEMDAPFWCNGAACFFEGIDVHWKENGTLVQVATISGNMF 180
 121 MEWYELFQJGNCCTPHLRPEMDAPFWCNGAACFFEGIDVHWKENGTLVQVATISGNMF 180
 121 MEWYELFQJGNCCTPHLRPEMDAPFWCNGAACFFEGIDVHWKENGTLVQVATISGNMF 180
 181 NQAKWVQDNDEGTGYETWNVKASPEKGAETWFDSDSKVLRITNKLAEGFAEFKNI 240
 181 NQAKWVQDNDEGTGYETWNVKASPEKGAETWFDSDSKVLRITNKLAEGFAEFKNI 240
 181 NQAKWVQDNDEGTGYETWNVKASPEKGAETWFDSDSKVLRITNKLAEGFAEFKNI 240
 181 NQAKWVQDNDEGTGYETWNVKASPEKGAETWFDSDSKVLRITNKLAEGFAEFKNI 240
 241 ETTNTRIFLYSGEPYLGNETSVFGPTGNKTLGLAIKRFYFPKPHLPTEFLLSLQIF 300
 241 ETTNTRIFLYSGEPYLGNETSVFGPTGNKTLGLAIKRFYFPKPHLPTEFLLSLQIF 300
 301 DAVIVHKQFYLFYNFEYWFLLPMKFPFIKITYEELPLPINKTSLGL 346
 301 DAVIVHKQFYLFYNFEYWFLLPMKFPFIKITYEELPLPINKTSLGL 346

RESULT 3
 S-10-653-595-126

Sequence 126, Application US/10653595

Publication No. US20040048304A1

GENERAL INFORMATION:

APPLICANT: Ruben et. al.

TITLE OF INVENTION: 95 Human secreted proteins

FILE REFERENCE: P2027P1C1
 CURRENT APPLICATION NUMBER: US/10/653,595
 CURRENT FILING DATE: 2003-09-03
 PRIOR APPLICATION NUMBER: US 09/397945
 PRIOR FILING DATE: 1999-09-17
 PRIOR APPLICATION NUMBER: PCT/US99/05804
 PRIOR FILING DATE: 1999-03-18
 PRIOR APPLICATION NUMBER: 60/078,566
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/078,576
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/078,573
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/078,574
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/078,579
 PRIOR FILING DATE: 1998-03-19
 PRIOR APPLICATION NUMBER: 60/080,314
 PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/080,312
 PRIOR FILING DATE: 1998-04-01
 PRIOR APPLICATION NUMBER: 60/078,578
 PRIOR FILING DATE: 1998-03-19
 Remaining Prior Application data removed - See File Wrapper or PALM.
 NUMBER OF SEQ ID NOS: 470
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 126
 LENGTH: 346
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SITE
 LOCATION: (242)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (246)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-10-653-595-126

Query Match 99.4%; Score 1916; DB 12; Length 346;
 Best Local Similarity 99.4%; Pred. No. 7,36-180;

Matches 344; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 MRGAGAAARASWCMALALMLAVVPGSRVSGIPSRHWPVYKRFPRPDPYCOA 60
 1 MRGAGAAARASWCMALALMLAVVPGSRVSGIPSRHWPVYKRFPRPDPYCOA 60
 61 KYTFCPTGSPIPVMBEGDDIEVFRLQAPVWEFRYGDLLGLKIMHDAIGFRSTLTGKNT 120
 61 KYTFCPTGSPIPVMBEGDDIEVFRLQAPVWEFRYGDLLGLKIMHDAIGFRSTLTGKNT 120
 121 MEWYELFQJGNCCTPHLRPEMDAPFWCNGAACFFEGIDVHWKENGTLVQVATISGNMF 180
 121 MEWYELFQJGNCCTPHLRPEMDAPFWCNGAACFFEGIDVHWKENGTLVQVATISGNMF 180
 121 MEWYELFQJGNCCTPHLRPEMDAPFWCNGAACFFEGIDVHWKENGTLVQVATISGNMF 180
 121 MEWYELFQJGNCCTPHLRPEMDAPFWCNGAACFFEGIDVHWKENGTLVQVATISGNMF 180
 181 NQAKWVQDNDEGTGYETWNVKASPEKGAETWFDSDSKVLRITNKLAEGFAEFKNI 240
 181 NQAKWVQDNDEGTGYETWNVKASPEKGAETWFDSDSKVLRITNKLAEGFAEFKNI 240
 181 NQAKWVQDNDEGTGYETWNVKASPEKGAETWFDSDSKVLRITNKLAEGFAEFKNI 240
 181 NQAKWVQDNDEGTGYETWNVKASPEKGAETWFDSDSKVLRITNKLAEGFAEFKNI 240
 241 ETTNTRIFLYSGEPYLGNETSVFGPTGNKTLGLAIKRFYFPKPHLPTEFLLSLQIF 300
 241 ETTNTRIFLYSGEPYLGNETSVFGPTGNKTLGLAIKRFYFPKPHLPTEFLLSLQIF 300
 301 DAVIVHKQFYLFYNFEYWFLLPMKFPFIKITYEELPLPINKTSLGL 346
 301 DAVIVHKQFYLFYNFEYWFLLPMKFPFIKITYEELPLPINKTSLGL 346

RESULT 4
 US-09-397-945-126

Sequence 126, Application US/09397945

Publication No. US20030065139A1

GENERAL INFORMATION:

APPLICANT: Human Genome Sciences, Inc. et al.
TITLE OF INVENTION: 95 Human secreted proteins
FILE REFERENCE: P2027PL
CURRENT APPLICATION NUMBER: US/09/397,945
CURRENT FILING DATE: 1999-09-17

PRIOR APPLICATION NUMBER: PCT/US99/05804
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,581
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,577
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,563
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,313
PRIOR FILING DATE: 1998-04-01
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 126

LENGTH: 347

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (242)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

LOCATION: (246)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (347)

OTHER INFORMATION: Xaa equals stop translation

-09-397-945-126

Query Match 99.4%; Score 1916; DB 12; Length 347;

Best Local Similarity 99.4%; Pred. No. 7.3e-180;

Matches 344; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 MRGAGNARGRACWALALIMLAVVGVGSRVSGIPSRBRMPVYKRPDRPDDPCQA 60
2 MRGAGNARGRACWALALIMLAVVGVGSRVSGIPSRBRMPVYKRPDRPDDPCQA 60
1 KYTFCFTGSPHPVMEGDDIEVERLQAPVWEFKYKGLGLKIMDAIGRSLTIGKNT 120
61 KYTFCFTGSPHPVMEGDDIEVERLQAPVWEFKYKGLGLKIMDAIGRSLTIGKNT 120
121 MEWEYELFOGNCCTFPHLRPMMDAPFWNCGAACPFGGIDVHWKENGTLVQVATISGNMF 180
121 MEWEYELFOGNCCTFPHLRPMMDAPFWNCGAACPFGGIDVHWKENGTLVQVATISGNMF 180
181 NQAKAVKQDNETGTYETWNVASPEKGAETWDSYDCKFPLRTFNKLAERGAEFKNI 240
181 NQAKAVKQDNETGTYETWNVASPEKGAETWDSYDCKFPLRTFNKLAERGAEFKNI 240
241 ENNYTRIFLYSGHPTVLTGNETSVFSGTGNKTLGLAIKRFYEPKPLPTKEFLISLQIF 300
241 ENNYTRIFLYSGHPTVLTGNETSVFSGTGNKTLGLAIKRFYEPKPLPTKEFLISLQIF 300
241 ENNYTRIFLYSGHPTVLTGNETSVFSGTGNKTLGLAIKRFYEPKPLPTKEFLISLQIF 300

QY 301 DAVIHKQFYLFYNEFWELPMKPEFIKITVEEILPLIRNNTLSGL 346
DB 301 DAVIHKQFYLFYNEFWELPMKPEFIKITVEEILPLIRNNTLSGL 346

RESULT 5
US-09-864-761-39378Sequence 39378, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Acemica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

PRIOR FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 39378

LENGTH: 163

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC001226.1

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6

OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 4.2

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3

OTHER INFORMATION: EST_HUMAN HT: H01255.1, EVALUATE 5.00e-59

OTHER INFORMATION: SWISSPROT HIT: O75503, EVALUE 2.00e-96
US-09-864-761-39378

Query Match 46.0%; Score 886; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 6.5e-79;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 184 AKWVKQNEGTGYETWNNVAKASPEKGAETWFDSDCKFVLRTEFNKLAERGAEFKNIEYN 243
1 AKWVKQNEGTGYETWNNVAKASPEKGAETWFDSDCKFVLRTEFNKLAERGAEFKNIEYN 60

Y 244 YTHFLVSGPPTVGNSTVFGPTGKTLGLAIKRFYFPKPHLPTEKFLSLQIFDAV 303
b 61 YTHFLVSGPPTVGNSTVFGPTGKTLGLAIKRFYFPKPHLPTEKFLSLQIFDAV 120

Y 304 IVHKQFYLFFNFYFWFLPMKFPFKITTYEELPIPIRNKTLISGL 346
b 122 IVHKQFYLFFNFYFWFLPMKFPFKITTYEELPIPIRNKTLISGL 163

RESULT 6
S-10-264-049-4034

Sequence 4034, Application US/10264049
Publication No. US20040005579A1
GENERAL INFORMATION:

APPLICANT: Birse et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PA133P1

CURRENT APPLICATION NUMBER: US/10/264,049

PRIOR FILING DATE: 2002-10-04

PRIOR APPLICATION NUMBER: PCT/US01/18569

PRIOR FILING DATE: 2001-06-07

PRIOR APPLICATION NUMBER: US 60/209,467

PRIOR FILING DATE: 2000-06-07

NUMBER OF SEQ ID NOS: 4360

SOFTWARE: PatentIn Ver. 3.1

SEQ ID NO 4034

LENGTH: 89

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (58)_"

OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids

S-10-264-049-4034

Query Match 24.5%; Score 472; DB 15; Length 89;
Best Local Similarity 98.9%; Pred. No. 1.6e-38;
Matches 88; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

258 GNETSVFGPTGKTLGLAIKRFYFPKPHLPTEKFLSLQIFDAVIVHKQFYLFFNFY 317
1 GNETSVFGPTGKTLGLAIKRFYFPKPHLPTEKFLSLQIFDAVIVHKQFYLFFNFY 60

318 WFLPKFPFKITTYEELPIPIRNKTLISGL 346
61 WFLPKFPFKITTYEELPIPIRNKTLISGL 89

SUITE 7
S-10-369-493-12714

Sequence 12714, Application US/10369493
Publication No. US2003023675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 12714
LENGTH: 328
TYPE: PRT
ORGANISM: Aspergillus nidulans
US-10-369-493-12714

Query Match 5.1%; Score 97.5; DB 15; Length 328;
Best Local Similarity 26.1%; Pred. No. 0.69;
Matches 41; Conservative 19; Mismatches 50; Indels 47; Gaps 7;

QY 168 TLVQVATLISGNMFQNA-----KWKQDNETGYETWNNVAKASPEKGAET 212
DB 57 TLVSGFQSNHTQVAAVARKLGLVALVQEKWVMDKNGVYDRAKNIQISRLMGADT 116

QY 213 WFDSDCKFVLRTEFNKLAERGAET-----FKNIETNTRIFLYSGEPTVGNSTVFGPTG 268
DB 117 RLD-----MSGFIEHKLTKNLE---EETIRARQGRKYI-----PAGASD 154

QY 269 NKTGLAIKRFYFPKPHLPTEKFLSLQIFDAVIV 305
DB 155 HPLGGLGFAR--WAFVEVQKE-----LGVFDTVIIV 185

RESULT 8
US-10-369-493-3172

Sequence 3172, Application US/10369493
Publication No. US2003023675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 3172

LENGTH: 290

TYPE: PRT

ORGANISM: Neurospora crassa

US-10-369-493-3172

Query Match 5.0%; Score 96; DB 15; Length 290;
Best Local Similarity 22.8%; Pred. No. 0.82;
Matches 69; Conservative 40; Mismatches 94; Indels 100; Gaps 19;

QY 5 AGAARGASWCMALALMLAVVPGMSRVSGIPSR-----RHMPPYKRF--DFRKPDPYC 58
DB 6 AGAALALASGVSAERYI-----GFNSGNITIPSRKAKFKKDWQVEFTTAQLKNSPGVEN 59

QY 59 QAK-YT-----FCPTGSPIPYWE-----GDDI--EYFRLOAPWEEFYK 95
DB 60 AVRLYTNTICAVSQRSEIEAFEALETNTKILGWMASGNTTIEPEIKALONGI--AKYG 117

QY 96 DLGLHKLIMDAIG--FSTLTG-----KNYTWETLFO--LGNCTF 134
DB 118 KKLTLDT--IIGASIGSEDLVRSVTVGIQNSGVAGPAELVYFIADWKKARQGTAIANVPI 176

QY 135 PHL-----REPMDAPFMCNGCAACFFBSIDVHMKKENTLYOVATISGMENOM 183
DB 177 GHVDYTDAMTNGNKKPVYIDAVDW-----GVDDYFYEEKGNKNIN--SGYLDRA 226

QY 184 AK-----WVKQDNETGYETWNNVAKASPEKGAETWFDSDCKFVLRTEFNKL 230


```
227 YDAEGAVGKPIWV--TETGMPVGTWDAATINQOYMGVEGCKRL-----FGKV 275
231 AEF 233
280 PTF 282

;SULT 9
;10-408-765A-58
Sequence 58, Application US/10408765A
Publication No. US20040101874A1
GENERAL INFORMATION:
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Fahy, Eoin D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Glenn, Gary W.
APPLICANT: Warnock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
FILE REFERENCE: 660088.465
CURRENT APPLICATION NUMBER: US/10/408,765A
CURRENT FILING DATE: 2003-04-04
NUMBER OF SEQ ID NOS: 3077
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 58
LENGTH: 836
TYPE: PRT
ORGANISM: Homo sapiens
;10-408-765A-58

Query Match 5.0%; Score 95.5; DB 16; Length 836;
Best Local Similarity 20.5%; Pred. No. 3.8; Indels 73; Gaps 12;
Matches 61; Conservative 39; Mismatches 125;

67 TGSPRPV-----MEGDDIEVFRLOAPVWEFKYGDLLG---HLKIMDAIGRSTLTG 116
272 TSAASPAGVGTGMDEDDPS-----KNPFYSSPSSPDKLTITIKTLRARESDDE 324
117 KNYIMWEYLFQNGCTPPLRPEMDAPFCNQGACFEGIDVHKENG-TLVQVATI 175
325 TQIKNMWEBAAGLLT--SLKDNVP-----KATGVLYDVNMKYHMTGLTLREVSCK 376
176 SGNMENQAKWKODNETGIYETWNKASPEKGAETPDSYDSKFLRFPNKLAEGBA 235
377 LRRLIQDHAENVQ---GAIREIDIDIERFOKGSGTGYQ----- 415
236 EFKNIETN-YRRIPLYSGEPTYLGNETSVFGPTGKTLGLAIK-----RF 279
416 EMKQKQKULYQELLTQEGQASFOGKONVFGIVRVVQEFMKVKHLIDSLIDFLNPRF 475
280 YYPEKPHLPTEKFLSLIQIPDAVIVHKQFY-----LFVNFEEWFLMKRPFK 328
476 QPFGKPGIYTRERELQTF-IREVGTVLSQVSKVHNGSEILFSFYODLVITLPELRK 532

;SULT 10
;10-282-122A-58749
Sequence 58749, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: E11TRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58749
; LENGTH: 1238
; TYPE: PRT
; ORGANISM: Helicobacter pylori
;10-282-122A-58749

Query Match 4.9%; Score 94; DB 12; Length 1238;
Best Local Similarity 23.6%; Pred. No. 9.2;
Matches 62; Conservative 28; Mismatches 77; Indels 96; Gaps 15;

QY 110 FRSLTGNKYMWEYLFQNGCTPPLRPEMDAPFCN-----QGACFEGIDVHK 164
Db 513 FTQYVGKNSALVF-----NATTP-----WANGSIPKSNSTVRGGYEGVNWG 555
QY 165 ENGLVQVAT-----ISNMENQAKWKODNETGIYETWNVVASPEKGAETPDSYDC 219
Db 556 KTVYITGTFPADRYIYIGNMV-----TG-----NGAQTGGA--T 588
QY 220 SKYVLRFPNKLAEKAPKNIET--NYTRIPLYSGEPTYLGNETSVFGPTGKTLGLAI 276
Db 589 LNFVGAETEINIA--GATFKMLKITSONSYMTFMALGD-----SSGSARKINVSQ 634
QY 277 KRFY-----YPEKPHLPTEKFLSLIQIPDAVIVHKQFYF-----YNFEYW-FLPMK 323
Db 635 SDFYDMTGGGYDFGN-----GVFDSVFNKAYVFFQGTGTSYNKKNINFLAGN 683
QY 324 FPFY-KITYEIRPPIRNKTLG 345
Db 684 FKFQKTTIEKSVLSDASTYFEDG 706

RESULT 11
;10-282-122A-55615
; Sequence 55615, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
```

RESULT 12
5-10-424-599-262855
 Sequence 262855, Application US/10424599
 Publication No. US20040031072A1
GENERAL INFORMATION:
 APPLICANT: La Rosa Thomas J
 APPLICANT: Kovalic David K
 APPLICANT: Zhou Yihua
 APPLICANT: Cao Yongwei
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21 (53223)B

QY 4 GAGAGRRASMCWALLMLTAVRG-----MSRVSGLFRRHHMPVEXR---FDDRPK 53
Db 9 GHQLDSRFS-----IVEGEFTYTLMADEGF---HMYNGHETSFAVRER 53
QY 54 PDPRCAKATFCP-----TSPPIRVMSGDDIIEVFLQAPRWMEKYGDLGLHKIMD 106
Db 54 SEPLVAEVMFSGDELLSFPLANEPLVPSSEIDENEAVALPAPLPKQFTLL----- 105
QY 107 AIGRRSTLTGKNY-----TMEWYELFQDNC-----TFPIRLBEMDAPFMCNOGAAC 153
Db 106 -VGAFS--TANNKRRMALRTTMOQEAIVSGVVAFPGTLKDNQVNNELV---REAQ 159

```

Y      154 FPEGIDVHKNGLTVQVATISGNMENC--AKNV-----KONE 192
b      160 LYGDIDQMPVDYITLTITISIGFKIVPAKXIMKTDDAFVRIDEVISLKSNS 219
Y      193 TGIYETWVWVKA--EKA-----ETW-FDSY-----DOSKFLRFENL 230
b      220 NGLLYGLISQSPHDKSKMFKISKKEPFEDMPPMAGPGVIRDAKFAVQCHOEL 279
Y      231 AERGAFFKNIETNYTRIFLYSGEPTVIGNETSVFGPTGNKTLGLAKRFY 280
b      280 T-----LQFLKEDVAMGLMIQYKNSGGQVNIYVDREY 314

```

RESULT 14

3-09-374-046A-160

Sequence 160, Application US/09374046A

Publication No. US20030096951A1

GENERAL INFORMATION:

```

APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallee, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Agostino, Michael J.
APPLICANT: Steindinger II, Robert J.
APPLICANT: Spaulding, Vikki
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
FILE REFERENCE: GI 6075-83A
CURRENT APPLICATION NUMBER: US/09/374,046A
CURRENT FILING DATE: 1999-08-13
NUMBER OF SEQ ID NOS: 240
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 160
LENGTH: 742
TYPE: PRT
ORGANISM: Homo sapiens
-09-374-046A-160

```

```

Query Match      4.7%, Score 91; DB 10; Length 742;
Best Local Similarity 22.1%; Pred. No. 9.1;
Matches 62; Conservative 39; Mismatches 105; Indels 74; Gaps 16;

```

```

116 GKNTMEYELFQGNCTFPH-----LRPMQAP-FWCNQG---AACFPEGIDVHWK 164
325 GSAVQFHSWVFLV-VCAFPVFAIGALTTPQESPREFLENGKDEAMWVLKQVHDNMR 383
165 ENGLVQVATISG-MNFMQAKWKQDNENGIYETWVWVKA--SPKGAETVDSYCSKFV 223
384 AKGHPREVSVTHIKTHQDEDELIEIOSDTGTWYQWGVRL-SLGGQVGNFLSC----- 438
224 LRTEFKLAERGAEPKNIETN-----YTRIFLYSGEPTV-----GNETSVE-- 264
439 -----FGPEYRRLITLMMGVMFTMSFSYGLTWFPDMIRHLQAVDYASRTKVPFG 489
265 -----GPTGNKTLGLAKRFYPPKP-----HLPTKEFLSLIQ--IFDAVIHKOQFY--- 310
490 ERVEHVTENFTLENOJHRGGQYFNDKFIGRLKSVSFEDSLFEBCYFEDVTSNTFFRNC 549
311 -----LFYN---FEYWFLLPMKFPFIKITV-----EELPLPI 338
550 TFINTVFYNTDLEPKYKFNKR--LINSTFLANKEGCPLDV 587

```

SOUT 15
-10-616-263-160
Sequence 160, Application US/10616263
Publication No. US20040038276A1

```

/ GENERAL INFORMATION:
/ APPLICANT: Jacobs, Kenneth
/ APPLICANT: McCoy, John M.
/ APPLICANT: Lavallee, Edward R.
/ APPLICANT: Collins-Racie, Lisa A.
/ APPLICANT: Evans, Cheryl
/ APPLICANT: Merberg, David
/ APPLICANT: Treacy, Maurice
/ APPLICANT: Agostino, Michael J.
/ APPLICANT: Steindinger II, Robert J.
/ APPLICANT: Spaulding, Vikki
/ APPLICANT: Wong, Gordon G.
/ APPLICANT: Clark, Hilary
/ APPLICANT: Fechtel, Kim
/ APPLICANT: Genetics Institute, Inc.
/ TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
/ FILE REFERENCE: 00766.000103.5
/ CURRENT APPLICATION NUMBER: US/10/616,263
/ CURRENT FILING DATE: 2003-07-08
/ NUMBER OF SEQ ID NOS: 240
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 160
/ LENGTH: 742
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-616-263-160

```

```

Query Match      4.7%, Score 91; DB 12; Length 742;
Best Local Similarity 22.1%; Pred. No. 9.1;
Matches 62; Conservative 39; Mismatches 105; Indels 74; Gaps 16;

```

```

QY      116 GKNTMEYELFQGNCTFPH-----LRPMQAP-FWCNQG---AACFPEGIDVHWK 164
Db      325 GSAVQFHSWVFLV-VCAFPVFAIGALTTPQESPREFLENGKDEAMWVLKQVHDNMR 383
QY      165 ENGLVQVATISG-MNFMQAKWKQDNENGIYETWVWVKA--SPKGAETVDSYCSKFV 223
Db      384 AKGHPREVSVTHIKTHQDEDELIEIOSDTGTWYQWGVRL-SLGGQVGNFLSC----- 438
QY      224 LRTEFKLAERGAEPKNIETN-----YTRIFLYSGEPTV-----GNETSVE-- 264
Db      439 -----FGPEYRRLITLMMGVMFTMSFSYGLTWFPDMIRHLQAVDYASRTKVPFG 489
QY      265 -----GPTGNKTLGLAKRFYPPKP-----HLPTKEFLSLIQ--IFDAVIHKOQFY--- 310
Db      490 ERVEHVTENFTLENOJHRGGQYFNDKFIGRLKSVSFEDSLFEBCYFEDVTSNTFFRNC 549
QY      311 -----LFYN---FEYWFLLPMKFPFIKITV-----EELPLPI 338
Db      550 TFINTVFYNTDLEPKYKFNKR--LINSTFLANKEGCPLDV 587

```

Search completed: June 8, 2004, 14:11:56
Job time : 72.1222 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

Run on: June 8, 2004, 13:59:38 ; Search time 24.9889 Seconds
(without alignments)
1331.883 Million cell updates/sec

Title: US-10-010-050A-2

Perfect score: 1927
Sequence: 1 MRGAGAGARASWCWALAL.....IKITYEIRPLPINKTISGL 346

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

searched: 283366 seqs, 96191526 residues

total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase:

1: PIR 78:***
2: PIR1:***
3: PIR2:***
4: PIR3:***
5: PIR4:***

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101.5	5.3	3036	2 T18995	hypothetical prote
2	95.5	5.0	359	2 A96816	P9K20.25 (imported
3	94	4.9	362	2 T33904	hypothetical prote
4	94	4.9	1238	2 A64596	hypothetical prote
5	93	4.8	561	2 T19604	hypothetical prote
6	93	4.8	1327	2 B90674	hypothetical prote
7	93	4.8	1349	2 E85524	AIDA-1 adhesin-lik
8	92	4.8	502	2 F70316	probable beta-bar
9	91	4.7	291	2 T21702	conserved hypotet
10	91	4.7	508	2 D85056	hypothetical prote
11	90	4.7	585	1 UCI486	probable aspartic
12	89	4.6	1353	2 T27404	neopululanase (EC
13	88.5	4.6	465	2 T51094	hypothetical prote
14	88.5	4.6	752	2 T51094	acid phosphatase (
15	88.5	4.6	1462	2 DJHUC6	hypothetical prote
16	88	4.6	411	2 D69459	DNA-directed DNA p
17	88	4.6	742	2 S27263	conserved hypotet
18	89	4.6	3194	2 D71917	synaptic vesicle p
19	87.5	4.5	672	1 A41043	toxin-like outer m
20	87.5	4.5	755	2 T41912	acetate-CoA ligase
21	87.5	4.5	1162	2 T37889	structural phospho
22	87	4.5	619	2 T27026	probable nuclear p
23	86.5	4.5	388	2 T49253	hypothetical prote
24	86	4.5	411	2 AF1352	purple acid phosph
25	86	4.5	1086	2 JC6079	probable phosphoe
26	85.5	4.4	5035	1 I46646	chitin synthase (E
27	84.5	4.4	464	1 B59200	ryanodine receptor
28	84.5	4.4	465	1 T51095	acid phosphatase (
29	84	4.4	247	2 H69493	phosphoesterase-re

30	83.5	4.3	441	2 E71425	hypothetical prote
31	83.5	4.3	640	1 A30452	utromodulin precurs
32	83.5	4.3	984	2 T44496	cellulose 1,4-beta
33	83	4.3	228	2 H69399	hypothetical prote
34	83	4.3	552	2 E72283	alpha-galactosidas
35	83	4.3	620	2 T27008	hypothetical prote
36	83	4.3	653	2 E70383	organic solvent to
37	83	4.3	1086	2 T33853	hypothetical prote
38	83	4.3	4563	1 LEHUB	apolipoprotein B-1
39	82.5	4.3	403	2 AF3413	fosmidomycin resis
40	82.5	4.3	509	2 S49349	cyprosin (EC 3.4.2
41	82.5	4.3	704	2 B83914	beta-xylosidase (1
42	82	4.3	386	2 T02640	hypothetical prote
43	82	4.3	456	2 AH0842	probable glycopor
44	82	4.3	477	1 A6709	alpha-amylase (EC
45	82	4.3	553	2 T03858	hypothetical prote

ALIGNMENTS

RESULT 1

T18995

hypothetical protein C06B8.7 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T18995

R/Steward, C.

submitted to the EMBL Data Library, November 1996

A/Reference number: Z19057

A/Accession: T18995

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-3036 <WIL>

A/Cross-references: EMBL:Z81463; PIDN:CAB03852.1; GSPDB:GN00023; CESP:C06B8.7

A/Experimental source: clone C06B8

C/Genetics:

A/Genes: CESP:C06B8.7

A/Map position: 5

A/Intons: 102/3; 157/2; 288/2; 439/3; 542/1; 591/3; 798/3; 835/3; 1139/1; 1204/3; 1325/1

Query Match 5.3%; Score 101.5; DB 2; Length 3036;
Best Local Similarity 22.2%; Pred. No. 9.2; Mismatches 143; Indels 165; Gaps 30;

Matches	103;	Conservative	52;	Mismatches	143;	Indels	165;	Gaps	30;
QY	10	GRASNCWALALLMLAVVPWGR	---	VSGIPSRHWPVPYKRFDRPRKDDPYCOAKYTF	-	64			
DB	1451	GRLSNRWNNKIMLQKV	---	NFTRNSEAVWMLKSPQAVVP	---	1500			
QY	65	-CP-TGSPFIPWEGDDDT	---	EVFRLOAPVWEKYG	---	98			
DB	1501	NCSTVDNSGPIIESHRDLVYASANVFHW	---	ILMSNTFANNSRSGIAVALPDYDILAKQTH	---	1558			
QY	99	-----GHLKIMHDAIGFRSITLTGNYMWEMLFOLANGCTRP	---	-----HURPE	---	140			
DB	1559	SFWLTERPFRNSFFKILLDGYVAFANISSNNFTLN	---	-----TAPKQFGWELR-G	---	1607			
QY	141	MDAPFWCNOGAACPFEGIDVHM	---	-KENGCT	---	195			
DB	1608	MEKVLICERNRFFPNWG	---	---HMMIKLDATSQYLRQIDVESYVOYNYTER-NRRLNQRGD	---	1662			
QY	196	YIEFWNVKASPEKALFWFSDYCSKFLVLTENKLAFF	---	-GAEPKNI-ETNYT	---	245			
DB	1663	YVDMW	---	FRSYALGVFSQKVEVHFNFNELLIDFEIVSGAKYSDVFEETMTNMMW	---	1717			
QY	246	-----RIFLYSG	---	-----EPTLYGNETSY	---	269			
DB	1718	GTGNEAVIAQVDFEDDMATFTPAEMTFRVVSNDLSINFWNMPWRDQQLANATYLAEGVH	---	1777					
QY	270	-----XTIGAIARF	---	-----YYPFKPHLPTEKEFL	---	307			
DB	1778	DLHGRVYEDKNLTLITERWYSPFYNYRPFPRPYRITDVTIMPATLYLIGNVEHVHPNV	---	1857					

308 QFYLYNF---EYWFLEPKFPFKIT---YE-EIPLPIRNK 341
 1838 RIIIVGNIVAGGEIV-QPIRFKPINTEYMQKGEIPTEYRK 1879

RESULT 2

196816

19K20.25 [imported] - Arabidopsis thaliana

Species: Arabidopsis thaliana (mouse-ear cress)

Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

Accession: A96816

Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziani, A.; M.; Rooney, T.; Rowley, D.; Sakano, H.; Salzbey, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 Reference number: A86141; MUID:21016719; PMID:11130712
 Accession: A96816
 Status: preliminary
 Molecule type: DNA
 Residues: 1-359 <STO>
 Cross-references: GB:AE005173; NID:g3834323; PIDN:AAC81039.1; GSPDB:GN0141
 GenBank: F9K20.25
 Map position: 1

Query Match 5 0%; Score 95.5; DB 2; Length 359;
 Best local similarity 21.5%; Pred. No. 2;
 Matches 53; Conservative 39; Mismatches 97; Indels 57; Gaps 12;

62 YTCPTGSDIPVWEGDDIEFRLQAPVVEFYKGDILGHLKIMHAIQFRSLTIGKNTM 121
 80 YEMQPTGCDIPRFNGDFLRKKGKILF---VGSLSN-----N 116
 122 EYELFQDGNCTFPHLRPEMDAPFCNQGACFF---EGIDVHWKENGTLVQVAT--- 174
 117 YWVSL---SCMLHAAPVAKYTFQINKGSLSTETPEVGI-SVNLKNGFLVDVSDKTR 177
 175 ---ISGNMFMQAKVKKQDNENGIYETNNVAKSEKAGETPDSYDCKFVLRFNKLA 231
 172 GLITLIDTISRGNQWLGSDVALFNTFHW---SHTGRKATW-DYFQTDGDKIVKEMNME 226
 227 AFKIALTTWMSKWDHNDIPSKTRIVYQGVSPVHLNGE---WGRKGTCLGRTVFWGSPS 283
 281 YPFKPH 286
 284 YPGRN 289

SUBT 3

3904

Potheoretical protein Y25C1A.7a - Caenorhabditis elegans

Species: Caenorhabditis elegans

Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

Accession: T33904

Submitted to the EMBL Data Library, February 1999

Description: The sequence of C. elegans cosmid Y25C1A.

Reference number: Z21437

Accession: T33904

Status: preliminary; translated from GB/EMBL/DBD

Molecule type: DNA

Residues: 1-362 <KAL>

Cross-references: EMBL:AF125459; PIDN:AA012836.1; GSPDB:GN00020; CESP:Y25C1A.7a

Experimental source: strain Bristol N2; clone Y25C1A

A:Gene: CESP:Y25C1A.7a
 A:Map position: 2
 A:Insertions: 7/1; 30/3; 79/3; 146/1; 231/1; 282/3; 323/3

Query Match 4.9%; Score 94; DB 2; Length 362;
 Best local similarity 21.7%; Pred. No. 2.7;
 Matches 48; Conservative 27; Mismatches 88; Indels 58; Gaps 9;

QY 103 IMEDALGFRSTLTGKNYTWMEYLF-----QIGNCTFPHLR-----PEMD 142
 DB 48 VFDSOSTSTTRKGNFFSEFYQQFFDVERTDQVIRKLINSVIFTHRYVIGDPLQPIEDLW 107
 QY 143 APFCNQAGACFPEGIDVHWKENGTLVQVATISGNMFMQAKVKKQDNENGIYETNNV 202
 DB 108 GPFVWSV-----TLVPAIGFNG---LAOFIRNDGAKGTYSDFRM 145
 QY 203 KASPEKAG---TWFDSDYDGS--KEVLRTEFKLAEPGAEPFNIEFNTRYFLYSGEPT 255
 DB 146 AKIPEVATEISIFRAQFPKKNCKLIDFSMKKPKKINNLCESPFKKSD-----FCHONSFS 200
 QY 256 YLGNENSVFGPFGNKTIGLAIR---FYYPFKPHLPTEKFL 293
 DB 201 FC-NSHGFLTNGNLKTGSFSEQKMFYTSFITSASTIIFL 240

RESULT 4

A64596

hypothetical protein HP0609 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C:Accession: A64596

R/Tomb, J.F.; White, O.; Kierlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.K.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpe, P.D.; Smith, H.O.; Fraser, C.N.
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467; PMID:9252185
 A:Accession: A64596
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1238 <TOM>
 A:Cross-references: GB:AE000575; GB:AE000511; NID:g2313730; PIDN:AA007677.1; PID:g2313733

Query Match 4.9%; Score 94; DB 2; Length 1238;
 Best local similarity 23.6%; Pred. No. 13;
 Matches 62; Conservative 28; Mismatches 77; Indels 96; Gaps 15;

QY 110 FRSTLTGKNYTWMEYLFQDGNCTFPHLRPEMDAPFCN---QGAACFPEG:DDVHWK 164
 DB 513 FTQYGGKNSALVF-----NATTP-----WANGSIPKSNSTVRFGEYGVNWG 555
 QY 165 ENGLVQVAT---ISGNMFMQAKVKKQDNENGIYETNNVAKSEKAGETPDSYDCK 219
 DB 556 KTYGITGTFTADRYVITGNM-----TG-----NGAQGTGGA--T 588
 QY 220 SKFYLRTEFNKLAEPGAEPFNIEFNTRYFLYSGEPTGKNTGLAI 276
 DB 589 LNFVGALEINIA--GATFKDLKTTSONSYTFMALGD-----SSGSKXINVSQ 634
 QY 277 KRFY-----YPPKPHLPTEKFLSLIQIPDAVIYHKOFLF-----YNEFYW-FLPMK 323
 DB 635 SDFYDWTGGGYDTFGN-----GVFDSVNMKAYIKFQGTENSTYFKNTNPLAGN 683
 QY 324 PPFIL-KTYEIRLPIPLRNKTLG 345
 DB 684 FKQGGKTTTEKSVLSDASYTFDG 706

RESULT 5

T19604

hypothetical protein C31C9.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

Query 147 CNOGAACTFEFGIDVHWKENGTLVQ-----VATISGNMFNMQAKWVKOD-NETGIYYE 198
Db 1083 SONGSDRHHVGMVAGYGNSDSKTISRTGYRAKASVYNVSGVLATWVADDSRRNGAYLD 1142
QY 199 TANNVKAPEKGAETWFD-----SYDCKFEVLRT---FNKLAEF 233
Db 1143 SW-----AQYSWFDNTVTKGDDIQSESYKSKGFTASLEAGYKHKLAEF 1184
RESULT 7
 E85524
 Probable beta-barrel outer membrane protein 20402 [imported] - Escherichia coli (strain C)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85524
R:Perma, N.T.: Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
 Miller, L.; Grobeck, E.U.; Davis, N.W.; Lam, A.; Dinalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85524
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1349 <STO>
A:Cross-references: GB:AE005174; NID:g12513130; PIDN:AAQS4657.1; GSPDB:GNC0145; UWGP:Z0402
C:Genetics:
A:Experimental source: strain 0157:H7, substrain EDL933
Query Match 4.8%; Score 93; DB 2; Length 1349;
Best Local Similarity 20.3%; Pred. No. 17;
Matches 59; Conservative 32; Mismatches 93; Indels 106; Gaps 15;
QY 1 MRBGGARAGRGASWQALALMLAVPVGKSRVSGIPSRHHVPPVRYRFPDRPKDPYQCA 60
Db 966 LARQGANSG-----NMVLTSGSDSPELQEP-----DPMPPDEPNPNP 1004
QY 61 KYTFCTGPSPIPVMEGDDIEVFLQAPWFEKYGDLGLKLI-----MDAIG-- 109
Db 1005 EPNNPPTPTPGDLNVDDL--RPEA-----GSYIANLAAMNTFTTRLHERLTNTY 1054
QY 110 FRSTLTG--KNYTM-----EWEY-----LFQIGNCTFHLRREMDFPFW 146
Db 1055 YTDWVTGQOKQYTMMRHEGGHNKWRDGSGLKTSNRVYLDLG-----DVAQW 1104
QY 147 CNOGACFEFGIDVHWKENGTLVQ-----VATISGNMFNMQAKWVKOD-NETGIYYE 198
Db 1105 SONGSDRHHVGMVAGYGNSDSKTISRTGYRAKASVYNVSGVLATWVADDSRRNGAYLD 1164
QY 199 TANNVKAPEKGAETWFD-----SYDCKFEVLRT---FNKLAEF 233
Db 1165 SW-----AQYSWFDNTVTKGDDIQSESYKSKGFTASLEAGYKHKLAEF 1206
RESULT 8
 F70316
 conserved hypothetical protein aq_175 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
C:Accession: F70316
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Over-
V.
Nature 392, 353-359, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: F70316
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-502 <AOQ>
A:Cross-references: GB:AE000677; NID:g2982900; PIDN:AAQ06534.1; PID:g2982916; GB:AE00065
C:Genetics:
A:Experimental source: strain VF5
C:Gene: aq_175

C:Superfamily: probable 60K inner membrane protein; stage III sporulation protein homolog
F:293-486/Domain: stage III sporulation protein homology <SPOR>

Query Match 4.8%; Score 92; DB 2; Length 502;
Best Local Similarity 21.3%; Pred. No. 6;
Matches 63; Conservative 51; Mismatches 114; Indels 68; Gaps 15;

```

QY 70 PIPVMEGDDIEFRLQAPVMEFKYGDLLGHKIMDAIGFSTLTGKNYTWEMWELFOL 129
Db 118 PLEIFGNDLDO-KINFEVEYELKEGK-NSVELHKEIKVKKILSYKNGAI----- 166
QY 130 GNCSTPHLRPE-MDAPFWCNQGA---ACFEGIDVHKENGTLVQVATISGMNENQMA 184
Db 167 -----HLSVEGLKRPFWFVSVSPDDKAFYHVGPV-IKINGEVRLDV----- 209
QY 165 KWKVKQDNENGIYYETWNNVAKSPKGAETWSDYSCKEVLTFTFNKLAFFGAEFKNIETNY 244
Db 210 -----DDKGINEFEGNIEFGSESRFFPKGAKOVQKHIVK---VLDGKFFVSLT-- 238
QY 245 TRIFLYSGEPT-YLG-----NENSVFGPTGKNKTLGLAIK--RFYYPFKPHLPTEKPL 233
Db 259 ---FLYDEKITYLAKAKYARLRGLVDTLDMGTLAKTIIVKPLFLYIYEH--TGSWV 313
QY 294 LSLT-----QTFDAVIHAKQFYLYFNPEYWFLEPKKPFYIKITYEELPIRANKTL 343
Db 314 LSLVLTFTVHIFLPPGLGKSVSMOKIQE--LAPKMKIKOKIKXDDPVKXQEEEM 367

```

RESULT 9

T21702
Hypothetical protein F33E2.5 - *Caenorhabditis elegans*

Species: *Caenorhabditis elegans*
Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
Accession: T21702

Legend, N.

Submitted to the EMBL Data Library, January 1997

Reference number: Z19461

Accession: T21702

Status: preliminary; translated from GB/EMBL/DBJ

Molecule type: DNA

Residues: 1-291 <KIL>
Cross-references: EMBL:Z84574; PIDN:CA806543.1; GSPDB:GN00019; CESP:F33E2.5

Experimental source: clone F33E2

Genetics:

Gene: CESP:F33E2.5

Map position: 1

Introns: 252/1

Query Match 4.7%; Score 91; DB 2; Length 291;
Best Local Similarity 17.9%; Pred. No. 3.7;
Matches 57; Conservative 47; Mismatches 92; Indels 122; Gaps 13;

```

Y 57 YCOAKYTFCD---TSSPIPVMEGDDIEFRL-QAPVMEFKYGDLLGHKIMDAIGFR 111
b 25 YPEARVPTLPKILKTAADFQKLGSKDRALFTLSQSPINDKCTETIKVLLBQ---PQ 80
Y 112 STLTKNY-----TWMEYELFOLGNCCTPHLRPEMDAPFWCNQGA 151
b 81 DTIDIPNAYTAVSMALNFKIGONCHDFOPLDWYKNIK-CYFTYKTKITEGVQAVYFNDENG 133
Y 152 ACFFEGIDVHKENGTLVQVATISGMNENQMAKWK--QDNMTGYYETWNNVAKSPENG 203
b 140 NQGLAGV-----CYATPAG-----KWKAKIQTEQTLIRFEP--QKEP--- 175
Y 210 AETWEPDYSCSFVLTENKLAFFGAEFKNIETNYTRIFLYSGEPTLTGKNYTWEMWELFOL 269
b 177 ---FPUNSKSRFPDPAFKKFAFYG---RYCETEGGRVLWVGKA----- 214
Y 270 KTLGAIKRFYFPPHLPTEKFLSLDQIFRAVIVHKOFYLYF-----NF 315
b 215 ---DGOIVHNVHVPVCMQWTKGCAQCAKHFF 243
Y 316 EYWFLEPKKPFYIKITYEE 333

```

Db 244 EYWMIAPEBPVYQSKKE 261

RESULT 10

Db5056

probable aspartic proteinase (imported) - *Arabidopsis thaliana*

Species: *Arabidopsis thaliana* (mouse-ear cress)
Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
Accession: DB5056

R:anonymus, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.

A:Reference number: A85001; MUID:20083488; PMID:10617198

Accession: DB5056

Status: preliminary

Molecule type: DNA

Residues: 1-508 <STO>

Cross-references: GB:NC_001268; NID:g7267203; PIDN:CA877914.1; GSPDB:GN00140

Gene: ATG04460

Map position: 4

Superfamily: oryzasin, saposin repeat homology

Query Match 4.7%; Score 91; DB 2; Length 508;
Best Local Similarity 23.8%; Pred. No. 7.5;
Matches 36; Conservative 17; Mismatches 48; Indels 50; Gaps 7;

```

QY 57 YCOAKY-----TFCTGSPIPVMEG-----DDIEFRLQAPVMEF----- 92
Db 125 YFHSKYKASQSSYKRGKXASIRYGTGALSIGFSNDVAKGDIWKEQETIATSEPGI 184
QY 93 -----KYGDLGHUKIMDAIGFRSTLTGKNYTWEMWELFOLGNCCTPHLRPEMDAPFWC 147
Db 185 TFLIAKFDGLG-----LGFKEISVG-NSIPVWYNNVEXG-----LVKEPIFFFWL 229
QY 148 N-----QGAACPFEGIDVHKENGTLVQY 172
Db 230 NRPKXPEGGEIYFGVDPKHFKEHTFVAV 260

```

RESULT 11

JCI486

neopullulanase (BC 3.2.1.135) - *Thermoactinomyces vulgaris*

N:Alternate names: alpha-amylase II

C:Species: *Thermoactinomyces vulgaris*

C:Date: 31-Dec-1993 #sequence_revision 18-Aug-1995 #text_change 16-Jun-2000

Accession: JCI486

R:Tonozuka, T.; Ohtsuka, M.; Mogi, S.; Sakai, H.; Ohra, T.; Sakano, Y.

Biochim. Biophys. Acta 57, 395-401, 1993

A:Title: A neopullulanase-type alpha-amylase gene from *Thermoactinomyces vulgaris* R-47.

A:Reference number: JCI486; MUID:93222535; PMID:7763540

Accession: JCI486

Molecule type: DNA

Residues: 1-585 <TON>

Cross-references: GB:D13178; NID:g91625; PIDN:BA02473.1; PID:g998125

Function:

A:Description: hydrolysis of alpha-(1->4)-glucosidic linkages of pullulan to produce pan

A:Pathway: pullulan degradation

A>Note: also has alpha-amylase activity

C:Superfamily: neopullulanase; alpha-amylase core homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:293-424/Domain: alpha-amylase core homology <AMY>

F:329,354,421/active site: Glu, Glu, Asp #status predicted

Query Match 4.7%; Score 90; DB 1; Length 585;
Best Local Similarity 21.5%; Pred. No. 11;
Matches 53; Conservative 30; Mismatches 68; Indels 96; Gaps 15;

```

QY 101 IKIMDAI-----GFRSTLT--GKNYTWEM-----YELFOLGNCCTF 134
Db 234 IKIIDAVERNAGDQFAFRDLQKGEOSRYKDWFIEDFPVSKTSRNYETFAVQVAM 293

```

SUTL 13
 1094
 Id phosphatase (EC 3.1.3.2) purple 1, precursor [similarity] - sweet potatcch
 Species: Ipomoea batatas (sweet potato)

```

Query Match      4.6%; Score 88.5; DB 2; Length 752;
Best Local Similarity 19.7%; Pred. No. 20;
Matches 67; Conservative 34; Mismatches 106; Indels 133; Gaps 18

20 LMT---AVPGMSRVSGIPSRHHMFVKRDFPRKDPYCOAKYTCPTSPIPVME 75
180 ILMGQYKVFVYVWSWVKGL---YHPTLDIRFDLISL-DE 216
76 GDDDIIVYRLQAPVMEKXYGDDLLGHKINMDALIGFSTLTGKX---YTMENYELFQLN 131
217 ---NAYVVD---LGFEGYSGNPATWVDVFFGMPDFED--- 247
132 STPPHLREPMADP---FWNCGAACEFEGIDDVYKENGNTLVQATISGNMFNMA-K 185
248 -----PADYETFGMHITMHNIGQLTV-----TFMKDETL-----ELTPPLAAYA 285

```


1b WKQDNETGIY-----YETWYKASPEKGAETWPDSDSCFVLTENKLAEPGAE 235
1b MAHNEDESTWYEVIRGVKAFDMNMNKTYDVAVDLFTIWRJARLSLPSMMITFE--- 342
1b 237 EKRIENNYRIEL-----YSGPPTLGNSTSVFGTGNKTGLAIK 277
1b 343 ---VDVNNQVQLTEEFNNOLLGGGGIYASGVSGKSLDELLOAFVSGG-TAGVVKI 393
1b 278 RFFYPPKPHLPTEKFLSL-LQIFDAVIVHKQFYLFYNEF 316
1b 399 KLYPPYAP-----ILSTFADPFTSVIPMK-YIFDWE 429

RESULT 15

MA-directed DNA polymerase (RC 2.7.7.7) alpha catalytic chain - human

Species: Homo sapiens (man)
Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 27-Oct-2003
Accession: S00257; A30440; I57513; S12665
Mong, S.W.; Wahl, A.F.; Yuan, P.M.; Arai, N.; Pearson, B.E.; Korn, D.; Hunt
MBO J. 7, 37-47, 1988
Title: Human DNA polymerase alpha gene expression is cell proliferation dependent and
Reference number: S00257; MIMD:88196030; PMID:3359994
Accession: S00257
Molecule type: mRNA
Residues: 1-1462 <WON>
Cross-references: EMBL:X06745; NID:g35567; PIDN:CAA2920.1; PID:g35568
Accession: A30440
Molecule type: protein
Residues: 438-449; 495-502; 'G', 838-848; 1090-1105; 1201-1216; 1397-1407; 1444-1453 <WON2>
Pearson, B.E.; Nashener, H.P.; Wang, T.S.
Mol. Cell. Biol. 11, 2081-2095, 1991
Title: Human DNA polymerase alpha gene: sequences controlling expression in cycling an
Reference number: I57513; MIMD:9112197; PMID:2005899
Accession: I57513
Status: translation not shown; translated from GB/EMBL/DDBT
Molecule type: DNA
Residues: 1-8 <RES>
Cross-references: GB:M64481; NID:g181617; PIDN:AAA52318.1; PID:g181618
Hsi, K.L.; Copeland, W.C.; Wang, T.S.F.
Nucleic Acids Res. 18, 6231-6237, 1990
Title: Human DNA polymerase alpha catalytic polypeptide binds ConA and RCA and contain
Reference number: S12665; MIMD:91057099; PMID:2243771
Accession: S12665
Molecule type: Protein
Residues: 19-37; 'C', 1406-1425; 'C' <HST>
Genetics:
Gene: GDB:POLA
Cross-references: GDB:120304; OMIM:312040
Map position: Xp22.3-Xp21.1
Superfamily: DNA polymerase
Keywords: DNA binding; DNA replication; nucleotidyltransferase; zinc finger
654-691/region: zinc finger CHCC motif
1249-1374/Region: zinc fingers

Query Match 4.6%; Score 88.5; DB 1; Length 1462;

Best local similarity 20.3%; Pred. No. 47;
Matches 71; Conservative 56; Mismatches 115; Indels 107; Gaps 20;

1b KRPDPKPPYCOAKYFCPTG-----SPFVMEGD---DIEV---FRL 85
1b 183 KKRSIGASPMFVSHTATAPSGKIASPVSRKEPPLTPVPLKRAFPAGDDVQVESTEEQ 242
1b 86 QAPWEEKYGDLLGLKIMDAIGFRSTLTGKNYTWEMWYLLFOLGNTCFPHLRPMDAF 145
1b 243 ESGAMEFEDGDF-----DEPMEVEV-----DLER-MAKA 272
1b 146 WCNQGAACFFEGIDVWKEN---GLVQVATISGNMFQAKM-VKQDNETGIYYETWN 201
1b 273 WDKES-----EPAEVYQGEADSGKGV---SYLGSFLPDVSCMDIDQEGDSFSVQEVQ 323
1b 202 VKASP---EKGAET-----WPDSTY---CSKFLVLTENKLAEPGAE-----FKNI 240

Db 324 VDSHLPVKGADSEQVHFHTWLDAYEDQYNQGVVFLFGKWMISAEETHVSCVMWKNI 383
Qy 241 ETNYTRIFLYSGEPPTLGNSTSVFGTGNKT-LGLAIRFYPPKPHLPTEKFLSLLOI 299
Db 384 ERT-----LY-----FLREMKIDANTGKETGTPISMKDVVEEPDEKATKXKIK- 429
Qy 300 FDAVIVHKQFYLFNFEYFLPMKFPFIKIY---EIPLRNKTLS 344
Db 430 FSKKPEKN---YAFELPDVEKSEYLEVKYSAMPOLPDLKGETFS 474

Search completed: June 8, 2004, 14:08:07
Job time : 25.9889 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: June 8, 2004, 13:56:38 ; Search time 14.4167 Seconds

(without alignments)
1249.684 Million cell updates/sec

file: US-10-010-050a-2
object score: 1927
sequence: 1 MRGAGARGRASWCMALAL.....IKITTEPIPIRKTKISGL 346

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

imum DB seq length: 0
imum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : SwissProt_42:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1927	100.0	407	1	CLN5_HUMAN
2	92	4.8	1705	1	OXNA_AQUAE
3	31.5	4.7	1705	1	PTPV_MOUSE
4	90	4.7	585	1	NER2_THERV
5	89.5	4.6	427	1	ARP3_ACA
6	88.5	4.6	1462	1	DPOA_HUMAN
7	88	4.6	411	1	TRUD_ARCFU
8	87.5	4.5	672	1	ACSA_METSO
9	87.5	4.5	755	1	PI00_HSVTJ
10	87	4.5	1184	1	CHSD_EMENT
11	85.5	4.4	5035	1	RPR1_PIG
12	85	4.4	254	1	SSA2_MOUSE
13	83.5	4.3	352	1	AB5C_MYCAV
14	83.5	4.3	564	1	TM16_HUMAN
15	83.5	4.3	640	1	URDM_HUMAN
16	83.5	4.3	2167	1	BPH1_YEAST
17	83	4.3	228	1	YC01_ARCFU
18	83	4.3	889	1	CI22_HUMAN
19	83	4.3	999	1	HGPB_HAEIN
20	83	4.3	1028	1	FDXG_HAEIN
21	83	4.3	1310	1	ACE_RABIT
22	83	4.3	4563	1	APB_HUMAN
23	82	4.3	2492	1	ATRX_HUMAN
24	82	4.3	2492	1	ATRX_PANTR
25	81.5	4.2	514	1	CHH2_SCHPO
26	81.5	4.2	606	1	VE1_HPV8
27	81.5	4.2	5037	1	RVR1_RABIT
28	81	4.2	644	1	URDM_RAT
29	81	4.2	1451	1	DPOA_RAT
30	80.5	4.2	282	1	HCHA_ECO57
31	80.5	4.2	544	1	OXNA_BORBU
32	80.5	4.2	555	1	POLG_DENIT
33	80.5	4.2	738	1	SEC6_DROME

34	80.5	4.2	835	1	VP3_ROTSL
35	80	4.2	429	1	CLN4_PHAVU
36	80	4.2	509	1	APR1_ORISA
37	80	4.2	2492	1	ATRX_PONPY
38	79.5	4.1	511	1	SYFA_MERTH
39	79.5	4.1	603	1	VE1_HPV8
40	79	4.1	508	1	ASPR_HORVU
41	79	4.1	682	1	TDR5_HUMAN
42	79	4.1	928	1	ODO1_RICCN
43	79	4.1	2292	1	POLG_EMCVB
44	79	4.1	2292	1	POLG_EMCVD
45	78.5	4.1	218	1	VLYS_BPXC

ALIGNMENTS

```

RESULT 1
CLN5_HUMAN
ID CLN5_HUMAN STANDARD; PRT; 407 AA.
AC 075503;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ceroid-lipofuscinosis neuronal protein 5 (CLN5 protein).
NM
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., VARIANT VLINCL ASN-279, AND VARIANT ARG-368.
RX TISSUE=Fetal brain;
RC MEDLINE=96324783; PubMed=9662406;
RA Savukoski M., Kloockars T., Holmberg V., Santavuori P., Landier E.S.,
RA Peltonen L.;
RT "CLN5, a novel gene encoding a putative transmembrane protein mutated
RT in Finnish variant late infantile neuronal ceroid lipofuscinosis.";
RL Nat. Genet. 19:286-288(1998).
RN [2]
RP SUBCELLULAR LOCATION, AND GLYCOSYLATION.
RX MEDLINE=21968572; PubMed=11971870;
RA Isoomppi J., Vesa J., Jalanko A., Peltonen L.;
RT "Lysosomal localization of the neuronal ceroid lipofuscinosis CLN5
RT protein.";
RL Hum. Mol. Genet. 11:885-891(2002).
CC
CC -!- FUNCTION: Not known.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- PTM: Glycosylated.
CC -!- DISEASE: Defects in CLN5 are the cause of Finnish variant late-
CC infantile neuronal ceroid lipofuscinosis (VLINCL) [MIM:256731];
CC also known as ceroid lipofuscinosis neuronal 5 (CLN5). VLINCL is a
CC fatal childhood neurodegenerative disease characterized by
CC progressive visual and mental decline, motor disturbance, epilepsy
CC and behavioral changes. The first symptom is motor clumsiness,
CC followed by progressive visual failure, mental and motor
CC deterioration and later by myoclonia and seizures.
CC
CC -!- DATABASE: NAME=NCU CLN5;
CC NOTE=Neural Ceroid Lipofuscinoses mutation db;
CC WWW="http://www.ucl.ac.uk/ncu/CLN5.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC
CC DR EMBL, AF068227; AAC27614.1; -.
CC DR Genew; HGNC:2076; CLN5.
CC MIM; 608102; -.

```

```

NM: 256731; -
GO: GO:0016021; C: integral to membrane; TAS.
GO: GO:0008151; P: cell growth and/or maintenance; TAS.
Transmembrane; Lysosome; Glycoprotein; Neuronal ceroid lipofuscinosis;
Disease mutation; Polymorphism; Epilepsy.
TRANSMEM 75
CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL)
CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL)
CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL)
CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL)
CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL)
CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL)
CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL)
CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL)
VARIANT 279 279 D -> N (in VLINCL)
VARIANT 368 368 /FTID=VAR_005137
VARIANT 407 AA; 46339 MW; 449702DIDCBFE4 CRC64;
SEQUENCE 407 AA; 46339 MW; 449702DIDCBFE4 CRC64;

Query Match 100.0%; Score 1927; DB 1; Length 407;
Best Local Similarity 100.0%; Pred. No. 2, 8e-157;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 MRRGAGARGRASNCMALALMLAVPGMSRVSGIPSRHMPVYKRPDPKPYCOA 6C
b 62 MRRGAGARGRASNCMALALMLAVPGMSRVSGIPSRHMPVYKRPDPKPYCOA 121
Y 61 KYTFCPTGSPLPVWESDDLEVPRLQAPVWEFKYGDILGLKIMHDAIGRSTLTGKNT 120
b 122 KYTFCPTGSPLPVWESDDLEVPRLQAPVWEFKYGDILGLKIMHDAIGRSTLTGKNT 181
Y 121 MEWELEFQNGCTFPHRPEMDAPFMCNOGAACFEFGIDDVHMKENGTLVQVATISGMF 180
b 182 MEWELEFQNGCTFPHRPEMDAPFMCNOGAACFEFGIDDVHMKENGTLVQVATISGMF 241
Y 181 NQAKKWKQONETGIYETWNVKASPEKGAETWEDSCSKFVLRTFNKLAEGAEFKNI 240
b 242 NQAKKWKQONETGIYETWNVKASPEKGAETWEDSCSKFVLRTFNKLAEGAEFKNI 301
Y 241 ETVNTRFLYXSGPTVGNSTVGPFGNTLGLAIRFYPRKPHLPTKEFLSLIQIT 300
b 302 ETVNTRFLYXSGPTVGNSTVGPFGNTLGLAIRFYPRKPHLPTKEFLSLIQIT 361
Y 301 DAVIVHKQFLFYNFEXWFLPMKFPFIKITYEELPIPIRKNTLSGL 346
b 362 DAVIVHKQFLFYNFEXWFLPMKFPFIKITYEELPIPIRKNTLSGL 407

RESULT 2
XAA AOUAE STANDARD; PRT; 502 AA.
D 06561;
T 15-MAR-2004 (Rel. 43, Created)
T 15-MAR-2004 (Rel. 43, Last sequence update)
T 15-MAR-2004 (Rel. 43, Last annotation update)
E Inner membrane protein oxaa.
N OXAA OR AQ 175.
S Aquifex aeolicus.
S Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
X NCBI_TaxID=63363;
P SEQUENCE FROM N.A.
C STRAIN=V5;
X MEDLINE=98196666; PubMed=9537320;
A Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
A Graham D.E., Overbeek R., Shead M.A., Keller M., Aufay M., Huber R.,
A Redman R.A., Short J.M., Olson G.J., Swanson R.V.;
T "The complete genome of the hyperthermophilic bacterium Aquifex
T aeolicus."
L Nature 392:353-358(1998).
I -I- FUNCTION: Required for the insertion of integral membrane proteins
I into the membrane. Probably plays an essential role in the

```

```

CC integration of proteins of the respiratory chain complexes.
CC Involved in integration of membrane proteins that insert
CC independently and independently of the Sec translocase complex (By
CC similarity).
CC -I- SUBUNIT: Specifically interacts with transmembrane segments of
CC nascent integral membrane proteins during membrane integration (By
CC similarity).
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -I- SIMILARITY: Belongs to the OXA1/oxa family. Subfamily 1.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE00677; AAC06534.1; -.
DR PIR; F70316; F70316.
DR HAMAP; MF_01810; -.
DR InterPro; IPR001708; 60kDa innermem.
DR Pfam; PF02096; 60KD IMP; 1.
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 12 34 POTENTIAL.
FT TRANSMEM 308 330 POTENTIAL.
FT TRANSMEM 378 400 POTENTIAL.
FT TRANSMEM 447 469 POTENTIAL.
SQ SEQUENCE 502 AA; 58224 MW; 6C24B8405F2570AD CRC64;

Query Match 4.8%; Score 92; DB 1; Length 502;
Best Local Similarity 21.3%; Pred. No. 2, 6;
Matches 63; Conservative 51; Mismatches 114; Indels 68; Gaps 15;

QY 70 PIPVMEGDDIEVRLQAPVWEFKYGDILGLKIMHDAIGRSTLTGKNTWMEYLFOL 129
Db 118 PLRIFGNDDLDQ-KINFEVEIKKQ--NSVELIKELKVKILSYKQAI----- 166
QY 130 GNCPTFHARE-MDAFPWNOGA---ACFEGIDDVHMKENGTLVQVATISGMENQMA 184
Db 167 -----HLSVEGILKPPFWFVGSPPDDEAFYTHVGPV-LKINGEAVRLDV----- 209
QY 185 KWKQONETGIYETWNVKASPEKGAETWEDSCSKFVLRTFNKLAEGAEFKNIETNY 244
Db 210 -----DOLKGINEFBSNIEFGGEBSRFFPKAKDYKHIVYK---VKLGDKFVSLST--- 258
QY 245 TRIFLYSGPT-YLG-----NETSVFGPTGNTLGLAIK--PRYPFPKPHLPTKEFL 293
Db 259 ---FLYDGEXTIYLGAQDYARLRELGLVDTLMDGTLKIVKFLFLVYIYEH--TGSWV 313
QY 294 LSLV-----QIPDAIVHKQFLFYNFEXWFLPMKFPFIKITYEELPIPIRKNTL 343
Db 314 LSLVLTFLVIRLFLPLGYKSVSMQKQE--LAPMEKIKOKYKDDPVKQGEEMV 367

RESULT 3
PTPV_MOUSE STANDARD; PRT; 1705 AA.
ID PTPV_MOUSE
AC P70289;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Receptor-type protein-tyrosine phosphatase V precursor (EC 3.1.3.48)
DE (Embryonic stem cell protein-tyrosine phosphatase) (ES cell
DE phosphatase).
GN PTPV OR ESP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```



```
RA Sakano Y.;
RT "Crystal structure of Thermoactinomyces vulgaris R-47 alpha-amylase II
RT (TVAII) hydrolyzing cyclodextrins and pullulan at 2.6-A resolution.";
RT J. Mol. Biol. 287:907-921(1999).
RN [3]
RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=22047855; Pubmed=12051850.
RT Kamitori S., Abe A., Ohkaki A., Kaji A., Tonozuka T., Sakano Y.;
RT "Crystal structures and structural comparison of Thermoactinomyces
RT vulgaris R-47 alpha-amylase 1 (TVAI) at 1.6 A resolution and
RT alpha-amylase 2 (TVAII) at 2.3 A resolution.";
RT J. Mol. Biol. 318:443-453(2002).
TC -!- FUNCTION: Hydrolyzes pullulan efficiently but only a small amount
TC of starch. Endohydrolysis of 1,4-alpha-glucosidic linkages in
TC pullulan to form panose. Cleaves also (1-6)-alpha-glucosidic
TC linkages to form maltotriose.
TC -!- CATALYTIC ACTIVITY: Hydrolysis of pullulan to panose (6-alpha-D-
TC glucosylmaltose).
TC -!- COFACTOR: Binds 1 calcium ion per subunit.
TC -!- SUBUNIT: Monomer.
TC -!- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
TC -----
TC This SWISS-PROT entry is copyright. It is produced through a collaborat:
TC between the Swiss Institute of Bioinformatics and the EMBL outstation
TC the European Bioinformatics Institute. There are no restrictions on its
TC use, by non-profit institutions as long as its content is in no way
TC modified and this statement is not removed. Usage by and for commercial
TC entities requires a license agreement (See http://www.isb-sib.ch/announce/
TC or send an email to license@isb-sib.ch).
TC -----
TC EMBL; D13178; BAA02473.1; -.
TC PIR; JCI486; JCI486.
TC PDB; 1BVZ; 02-MAR-99.
TC PDB; 1G1Y; 14-MAR-01.
TC PDB; 1JF5; 25-DEC-02.
TC PDB; 1JF6; 25-DEC-02.
TC PDB; 1JF7; 18-DEC-02.
TC PDB; 1JF8; 12-AUG-03.
TC PDB; 1JL8; 18-DEC-02.
TC InterPro; IPR006589; Alp amyl cat sub.
TC InterPro; IPR006047; Glyco amyl cat.
TC InterPro; IPR004185; Glyco_hydro_131g.
TC InterPro; IPR007110; Ig-like.
TC Pfam; PF00126; alpha-amylase; 1.
TC Pfam; PF02903; alpha-amylase_N; 1.
TC SMART; SM00642; Amy; 1.
TC Carbohydrate metabolism; Hydrolase; Calcium-binding;
TC 3D-structure.
TC ACT_SITE 325 325
TC ACT_SITE 354 354
TC ACT_SITE 421 421
TC METAL 143 143 CALCIUM.
TC METAL 145 145 CALCIUM (VIA CARBONYL OXYGEN).
TC METAL 148 148 CALCIUM.
TC METAL 149 149 CALCIUM.
TC METAL 169 169 CALCIUM (VIA CARBONYL OXYGEN).
TC METAL 171 171 CALCIUM.
TC HELIX 3 5
TC STRAND 7 7
TC TURN 12 14
TC STRAND 15 19
TC TURN 20 21
TC STRAND 22 30
TC TURN 31 32
TC STRAND 36 42
TC TURN 44 45
TC TURN 48 49
TC STRAND 53 62
TC STRAND 66 74
TC TURN 76 77
TC STRAND 80 87
TC STRAND 93 97
TC TURN 98 99
```

```
FT STRAND 100 102
FT HELIX 105 108
FT STRAND 111 113
FT TURN 119 120
FT HELIX 126 130
FT STRAND 133 136
FT HELIX 138 141
FT TURN 146 147
FT TURN 151 152
FT STRAND 153 153
FT TURN 158 159
FT TURN 164 165
FT STRAND 167 167
FT HELIX 172 185
FT TURN 186 186
FT STRAND 189 192
FT STRAND 196 196
FT STRAND 212 212
FT TURN 214 216
FT HELIX 219 230
FT TURN 231 233
FT STRAND 236 240
FT TURN 242 243
FT STRAND 244 244
FT TURN 247 248
FT HELIX 250 258
FT HELIX 259 261
FT TURN 263 264
FT HELIX 265 267
FT STRAND 268 268
FT STRAND 270 270
FT STRAND 283 283
FT STRAND 285 285
FT TURN 291 292
FT STRAND 294 296
FT TURN 298 299
FT HELIX 301 317
FT TURN 318 318
FT STRAND 321 324
FT TURN 325 325
FT HELIX 327 329
FT HELIX 332 345
FT TURN 347 348
FT STRAND 350 353
FT TURN 360 366
FT STRAND 370 373
FT HELIX 374 384
FT TURN 385 386
FT HELIX 391 402
FT TURN 403 404
FT HELIX 407 410
FT TURN 411 412
FT STRAND 414 416
FT TURN 420 421
FT HELIX 425 428
FT TURN 431 431
FT HELIX 443 443
FT TURN 444 445
FT STRAND 449 453
FT TURN 454 455
FT HELIX 456 458
FT TURN 459 459
FT TURN 465 468
FT HELIX 476 478
FT HELIX 481 496
FT TURN 498 498
FT HELIX 499 502
FT STRAND 504 511
FT TURN 512 515
FT STRAND 516 523
FT TURN 524 525
FT STRAND 526 533
FT STRAND 539 544
```

```

1 STRAND 553 555
2 STRAND 561 561
3 STRAND 564 565
4 STRAND 566 567
5 STRAND 568 573
6 STRAND 575 576
7 STRAND 578 582
8 SEQUENCE 585 AA; 67467 MW; E311813A05A7791A CRC64;

Query Match
Best Local Similarity 21.5%; Score 90; DB 1; Length 585;
Matches 53; Conservative 30; Mismatches 68; Indels 96; Gaps 15;

101 IKIMEDAI-----GFRSTLT--GKNYTWEM-----YELFOLGNCFTF 134
234 IKIILDAVNEHAGDQEFAPFDVLQKGQGRYKDWFIEDFPVSKSRINYEFAVQVPM 293
135 PHLR--PAMD-----AFPCNQGACFPBG-----IDVHKENGTLVQV---A 173
234 PRLRTEPEVKEKELPDVAFPMWEG--IDGMRLDVANEVDAFMREFRRLVKSINPDA 349
174 TISGMNENOMAKVAKODNETGI-----YYETMNVKAS-----PEKG 209
350 LIVGEIWDASGMLGDDPDSDVMYTLFRESVIRFPATGEIHAERPDALITRAMLYPQQA 409
210 AE-TV--PDSDYCSKFEVLRTEFNKLAFFGAFFKNIEITNYTRI-----FLYSGEP--TYLGNE 260
410 AAGLWNLDSHDTERTFLTSCGNEAKF-----RLAVLPQMTYLTGLPRLIYDDE 457
261 TSVFPGPT 267
458 IGMAGAT 464

RESULT 5
3P3 ACACA STANDARD; PRT; 427 AA.
ARP3 ACACA
P53490;
01-OCT-1996 (Rel. 34, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
15-JUL-1999 (Rel. 39, Last annotation update)
Actin-like protein 3.
ARP3.
Acanthamoeba castellanii (Amoeba).
Eukaryota; Acanthamoebidae; Acanthamoeba.
NCBI_TaxID=5755;
[1]
SEQUENCE FROM N.A.
STRAIN=ATCC 30010 / Neff;
MEDLINE=96017709; PubMed=7593166;
Kelleher J.F., Atkinson S.J., Pollard T.D.;
"Sequences, structural models, and cellular localization of the
actin-related proteins Arp2 and Arp3 from Acanthamoeba.";
J. Cell Biol. 131:385-397(1995).
-!- FUNCTION: Part of a complex implicated in the control of actin
polymerization in cells (By similarity).
-!- SUBUNIT: Belongs to a complex composed of ARP2, ARP3, P41-ARC,
P34-ARC, P21-ARC, P20-ARC and P16-ARC (By similarity).
-!- SIMILARITY: Belongs to the actin family. ARP3 subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; U29610; AAA93068.1;
InterPro: IPR004000; Actin_like.
Pfam: PF00022; actin; 1.
SMART: SM00268; ACTIN; 1.
PROSITE; PS01132; ACTINS_ACT_LIKE; 1.

```

```

KW Structural protein; Cytoskeleton.
SQ SEQUENCE 427 AA; 48635 MW; A53A0965B1E7EAF6 CRC64;

Query Match
Best Local Similarity 4.6%; Score 89.5; DB 1; Length 427;
Matches 64; Conservative 49; Mismatches 96; Indels 101; Gaps 17;

55 DEYCOAKYTFCP--TGSPFIPVME---GGDDLE---VFRLLQAVVMEFKGDLGLHKIMWD 106
44 DPCRRARWPCWMAKXNIADIDFISDBAYNSKYVQITMVF--RHQCV----- 91
107 ALGFRSTLTGKNYTL--MEWYELFOLGNCFTFPHLR-----PEMDAPFCNQAAC 153
92 -----ENWTHMEGF-----WEHCIFKYLRCPEPDHFLTLPEPLNAPNREYTAI 137
154 PFEIGIDVHKENGTLVQVATISGMFPNOMAKVAKODNETGIYETW----- 200
138 MFE-----TFNVPGLYIVQAVLALAAWSVTSQVTEKLTGLVVISGQGVTHVIPAVEGY 192
201 ---NVKASPEKGAR--TFE-----DSYCSKFEVLRTEFNKLA--EFGARFKN 239
193 VIGSISKIIPLAGRDITTFVQLRLERNEKIPPAETLEVAKAIKETFSVVCDDIYKEFKK 252
240 IETNYTRIFLYSGEPTVIGNETSVFPGYGNK--TLGLAIKFEYYP--FKPHLPKKEPL 294
253 YTEPDKWF-----KTYEGIES-----VGKKPYNDVGVGERLGEIPEIFSSDFLT 302
295 SLQGFIDAVI 304
303 PLPKVDETI 312

RESULT 6
DPOA HUMAN
ID DPOA HUMAN STANDARD; PRT; 1462 AA.
AC P09884;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA polymerase alpha catalytic subunit (EC 2.7.7.7).
GN POLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=88196090; PubMed=3359994;
Wong S.W., Wahl A.F., Yuan P.-M., Arai N., Pearson B.E., Arai K.,
Korn D., Hunkapiller M.W., Wang T.S.-F.;
"Human DNA polymerase alpha gene expression is cell proliferation
dependent and its primary structure is similar to both prokaryotic
RT and eukaryotic replicative DNA polymerases.";
EMBO J. 7:37-47(1988).
[2]
SEQUENCE OF 1-8 FROM N.A.
MEDLINE=91172197; PubMed=2005899;
Pearson B.E., Nashener H.P., Wang T.S.;
"Human DNA polymerase alpha gene: sequences controlling expression in
RT cycling and serum-stimulated cells.";
Mol. Cell. Biol. 11:2081-2095(1991).
-!- FUNCTION: Polymerase alpha in a complex with DNA primase is a
replicative polymerase.
-!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
+ [DNA] (N).
-!- SUBCELLULAR LOCATION: Nuclear.
-!- MISCELLANEOUS: In eukaryotes there are five DNA polymerases:
alpha, beta, gamma, delta, and epsilon which are responsible for
different reactions of DNA synthesis.
-!- SIMILARITY: Belongs to the DNA polymerase type-B family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X06745; CAA29920.1; --
EMBL; M64481; AAA52318.1; --
PIR; S00257; DJHUC.
Genew; HGNC; 9173; POLA.
GK; p09884; --
XIM; 312040; --
GO; GO:0005634; C:nucleus; NAS.
GO; GO:0003889; F:alpha DNA polymerase activity; NAS.
GO; GO:0006260; P:DNA replication; NAS.
InterPro; IPR006172; DNA_pol_B.
InterPro; IPR006134; DNA_pol_B_dom.
InterPro; IPR006133; DNA_pol_B_exo.
InterPro; IPR004578; Pol2.
Pfam; PF00136; DNA_pol_B; 1.
Pfam; PF03104; DNA_pol_B_exo; 1.
PRINTS; PR00106; DNAPOLB.
SMART; SMO0486; POLBc; 1.
TIGR; TIGR00592; pol2; 1.
PROSITE; PS00116; DNA_POLYMERASE_B; 1.
Transferrase; DNA-directed DNA polymerase; DNA replication;
DNA-binding; Nuclear protein.
DNA_BIND 650 715 POTENTIAL.
DNA_BIND 1245 1376
SEQUENCE 1462 AA; 165860 MW; 255270BOA0DB38BE CRC64;

Query Match 4.6%; Score 88.5; DB 1; Length 1462;
Best Local Similarity 20.3%; Pred. No. 20;
Matches 71; Conservative 56; Mismatches 115; Indels 107; Gaps 20;

46 KREDFRKPDPYCAKTPPTG-----SPIVMEGD--DIEV---FRL 85
183 KKRSIGASPMFVHTATVPAGKIASPVSRKPPPLPVLPKAEAGDVQVESTEEQ 242
86 QAPVWEKCYDGLGHLKIMDAIGFRSTLGNKNTYMEWYELFQIGNTFPHLPENDAPF 145
243 ESGAMEEDDDP-----DEPVEVEV-----DLRP-MAAKA 272
146 WCNQGAACFEGIDVHWKEN--GLVQVATISGMFNOMAKW-VKQDNETIYYETWN 201
273 WKES-----EPAEVYQGEADSGKGV---SYLGSLPLPVSCMDIDQEDSSFSVQEVQ 323
202 VKASP---EKGAT-----WFDSD--CSKFVLTENKLABFGAE-----FKNI 240
324 VDSHSLPLVVGADBEQVHFHYWMDAYEDQYNOGVVFLFGKVMIESAETHVSCCVWKNKI 383
241 ETMYTRFLVSGSEPTYGNFVSVPFGTKNT-LGLAKRFYFPKPHLPKKEFLSLLOI 299
384 ERT-----LY-----FIPRMKIDLTNGKGTGPIISKDVEEFDEKATKYIKM--- 429
300 FDAVIVKQFYLPYNEFYWFLPMKFPFKITYE---EIDLPTRNKTL 344
430 FKSPVPEK---YAFELPDVPEKSELEVKYSAEMWQLDQDKGERFS 474

RESULT 7
TRUD_ARCFU STANDARD; PRT; 411 AA.
NC 028596;
10-OCT-2003 (Rel. 42, Created)
10-OCT-2003 (Rel. 42, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Probable tRNA pseudouridine synthase D (EC 4.2.1.70) (Pseudouridylylate synthase) (Uracil hydrolase).
TRUD OR APL677.
NC Archaeoglobus fulgidus.
NC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
NC Archaeoglobaceae; Archaeoglobus.

NCBI TaxID=2234;
[1]
SEQUENCE FROM N.A.
SPRAIN-VC-16 / DSM 4304 / ATCC 49558;
MEDLINE=9809343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Arlrich P., Kaine B.P., Sykes S.M.,
RA Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Could be responsible for synthesis of pseudouridine from
CC uracil-13 in transfer RNAs (By similarity).
CC -1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -1- SIMILARITY: Belongs to the pseudouridine synthase trnd family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; AB000987; AAB89568.1; --
DR PIR; D69459; D69459.
DR TIGR; APL677; --
DR HAMAP; MF_01082; --; 1.
DR InterPro; IPR001656; UPF0024.
DR Pfam; PF01142; UPF0024; 1.
DR TIGR; TIGR00094; TIGR00094; 1.
DR PROSITE; PS01268; UPF0024; 1.
DR tRNA processing; Lyase; Complete proteome.
FW ACT SITE 81
SQ SEQUENCE 411 AA; 47745 MW; 10985250D96922E CRC64;

Query Match 4.6%; Score 88; DB 1; Length 411;
Best Local Similarity 20.3%; Pred. No. 4.4; Indels 74; Gaps 13;
Matches 55; Conservative 45; Mismatches 97;

74 MEGDDIEVFRLOAPYWEKYGDLGHLKIMDAIG-----FRSTLGNKNTYMEWYEL 126
38 LSGDEGFLIIRVEKKNW-----DTLNFARVLNAGLSQKRISFAGTKORBALTYGFESI 92
127 FQIGNTFPHLPENDAPFWCNQGAACFEGIDVHWKE-----NGTLVQVATISGN 178
93 YGVKK-----EELERVMKAKIEVIGYARRAQLDGLDN 128
179 MEMQAKWKQDNETIYYETWNVKAPEKGAETWF--DSYDCKFVLRTFNGLAEFGAE 236
129 FF-RIRVYGGROGE--IFQETRN--ELMEKGTNFGGLRFGSIRITHEVGKLI----- 178
237 FKNIEYNYTRIF-LVSGSEPTYLGN-----TSVFGPTGNKTGLAKRFYFPKPHLPK 290
179 ---LQNNYBEAFVYAKFEGENEVRKIRIETMETROAKIGRLPYLRYSRNLQK 235
291 -----EFLSLDQIFDAVIVK-QFYLF 312
236 LREKSEEBALSLPRLNLMKMPFVAYQSYIF 266

RESULT 8
ACSA_METSO STANDARD; PRT; 672 AA.
ID ACSA_METSO

```

Db      341 VYAPLLIGTSLMYGADYDPFGKMKXIQHKVTLVLTAFYAVMFWKQAE--WPKY 399
Qy      218 DSKRYVLTENKLAE-----FG-AEFOKNIET--NYTRIPLYS-----251
Db      400 DLS--LRLGSGVGPINDEAMMYREHIGRGELQIMDTWQTEGTFLNSPLPTPLPK 457
Qy      252 GEPT--LIGNSVSGPTGNKI-LG-----LAIKFYYPFKHLEPTKEPLSLQITF--D 301
Db      458 GSCTPEPLGYPDLSILDERBNEVPLSGGNVVALKP--YP-----SMLRAFWGD 503
Qy      302 AVIVHKQFYLEFNYFEWFLP 321
Db      504 KERFMKEYWQF-----YMDVP 519

RESULT 9
ID      P100 HSVJ7       STANDARD;          PRT;       755 AA.
AC      P52519;
DT      01-OCT-1996 (Rel. 34. Created)
DT      01-OCT-1996 (Rel. 34, last sequence update)
DT      16-OCT-2001 (Rel. 40, last annotation update)
DE      Large structural phosphoprotein homolog (Pp100).
GN      U11.
OS      Human herpesvirus (type 7 / strain J1) (HHV7).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Alphaherpesvirinae; Simplexvirus.
OX      NCBI_Taxid=57278;
RN      (1)
RP      SEQUENCE FROM N.A.
RA      Nicholas J.;
RL      Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC      -! SIMILARITY: TO THE LARGE STRUCTURAL PHOSPHOPROTEINS OF HSV-6 AND
CC      HCMV UL32.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      EMBL; U43400; AAC54672.1; -.
DR      PIR; T41912; T41912.
KW      Matrix protein; Phosphorylation
SQ      SEQUENCE 755 AA; 86580 MW; 4083744CCCF91DA CRC64;

Query Match 4.5%; Score 87.5; DB 1; Length 755;
Best Local Similarity 20.7%; Pred. No. 10;
Matches 58; Conservative 37; Mismatches 98; Indels 87; Gaps 15;

Qy      144 PF-WCNGSANC-----PREGIDD--VHKKN-GTIYQVATISGN-----MNQNAKVV 187
Db      7 PFAMISDEACXELSRFENISSLPVVDIRENPWILSQCIWKTGNSINNVTYVNNILW- 65
Qy      188 KQDNETGIYVETWNVKASPEKGAETWFDSDCKEV-----LRTENKLAE 232
Db      66 -----LYHQQLCKKKRPDY-EEWQELIKYQKILKDYLEQROMITDYSLSLSPKVG- 116
Qy      233 FGAEFKKIEINRYRI--FLYSGEPTLGN-----ETSVFG--PTGNKTLGLAIKRF 279
Db      117 FETEEFKNAVADKLKLSGFLRMGVTVAHADVNLITTEBEIGNLQKAKKNMISFTYQI 176
Qy      280 YPEFKPH-----LPTKEFLSLDQ-----IPDAIVHKQFY 310
Db      177 VDPWMNNGYVYVNNILLYIGNLLITLHSGMMMEKALNTINERKQAILIKALENNKVF 256
Qy      311 LFVNEFWFLPMK-----FPPIKITYEIDLPPIRNKTLSGI 346
Db      237 SIYSQILSLPLTSHRVTSFFKILTEDPQVITKSIETLHAL 276

```



```

RESULT 10
HSD_EMENT STANDARD; PRT; 1184 AA.
D CHSD_EMENT
C P78611; 000744;
T 15-JUL-1999 (Rel. 38, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Chitin synthase D (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl
transferase D) (Class-V chitin synthase D).
N CHSD OR CHSE.
C Emericella nidulans (Aspergillus nidulans).
C Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
C Eurotiales; Trichocomaceae; Emericella.
C NCBI_TaxID=162425;
[1]
P SEQUENCE FROM N.A.
C STRAIN=FSSC 89;
X MEDLINE=96285566; PubMed=8709948;
A Mochiyama T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
"the Aspergillus nidulans genes chsd and chsd encode chitin synthases
which have redundant functions in conidia formation.";
Mol. Genet. Genet. 251:442-450(1996).
[2]
P ERRATUM.
X MEDLINE=97188530; PubMed=9037115;
A Mochiyama T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
Mol. Genet. Genet. 253:520-528(1997).
[3]
P REVISIONS.
A Mochiyama T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[4]
P SEQUENCE FROM N.A.
C STRAIN=FSSC 4;
X MEDLINE=96406387; PubMed=8810520;
A Specht C.A., Liu Y., Robbins P.W., Bulawa C.E., Iartchouk N.,
Winter K.R., Rugg P.C., Rhodes J.C., Dodge C.L., Culp D.W.,
Borgia P.T.;
"The chsd and chse genes of Aspergillus nidulans and their roles in
chitin synthesis.";
Fungal Genet. Biol. 20:153-167(1996).
[1]
P FUNCTION: Plays a major role in cell wall biogenesis.
-!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + [(1,4)-(N-acetyl-
glucosaminyl)](N) = UDP + [(1,4)-(N-acetyl-beta-D-
glucosaminyl)](N+1).
-!- SUBCELLULAR LOCATION: Plasma membrane-bound.
-!- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS
V.
-----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
C EMBL; D83246; BA11866.2; -
C EMBL; U52362; AAA97482.1; -
C InterPro; IPR004835; Chitin_synth_fng.
C InterPro; IPR001199; Cyt_B5.
C Pfam; PF03142; Chitin_synth_2; 1.
C Pfam; PF00173; heme_1; 1.
C Transferrase; Glycosyltransferase; Transmembrane; Cell wall;
M Multigene family.
T TRANSMEM 221 241
T TRANSMEM 476 496
T TRANSMEM 1039 1059
T TRANSMEM 1073 1093
T TRANSMEM 1097 1117
T SEQUENCE 1184 AA; 133504 MM; D6157184F154ED55 CRC64;

```

```

Query Match 4.5%; Score 87; DB 1; Length 1184;
Best Local Similarity 20.5%; Pred. No. 20;
Matches 75; Conservative 44; Mismatches 132; Indels 114; Gaps 20;

QY 63 TRCPPTSPPIPWEGDDIVF---RLQAPVWEFKYDGLGHKTHDAIGERS---TLT 115
D 189 TYC---SVITFMAPFVLKFCGMPKQARSAMREKIG-LISILIMIAFAVGFITFAT 244
QY 116 -----GNNYMEWELEFQNGCPEPHR--PEM-----DAPF-WONGANC 153
D 245 VCGTPETRLKINIGSGWIFHGQAYDLTKSTHPAAAGIPDMTNVLYDPHKYGGDGSF 304
QY 154 FREGID-----DVHWKENGTLVQVATISGNFMQMAKVKQDNETIYEYETW 200
D 305 PFQEVVGAQKGLITRENSDIPITNSGDLAW--PFCIAFNQDGS--SPRTVTSYNGW 360
QY 201 --NVKASPEKGAE-----TWFDSDCKSFVLRITNKLAFGAE-----FXN 239
D 361 ACHTSGSARKSPFSLKNSGDVYFTWEDTYNTSR-----KLAVYSGNVLDLNLWFPD 413
QY 240 IETNYTRIF-----LYSGEPTYLIG--NETSVFSGPTGNKTLIG-LAIR 276
D 414 TQVNYPTKCDLRDNDIDIGVDLTYFTQGEDKQIKCSQIILKVGSIIDTDVGTIASQV 473
QY 279 FYYPFKPHLPYKEFLLSLQIPDAVIVHKQFYLFYNFEXWPLPMKPPFKITYERIPLP 338
D 474 VLY-----VSLIFILISIVVKPAFALPQ--WPLAPRAAQKTSWGAVDSKA 518
QY 339 RNKTL 343
D 519 RNOQT 523

RESULT 11
RYR1_PIG STANDARD; PRT; 5035 AA.
AC 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ryanodine receptor 1 (Skeletal muscle-type ryanodine receptor) (RYR1)
DE (RYR-1) (Skeletal muscle calcium release channel).
GN RYR1 OR CRC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Norwegian Landrace; TISSUE=Skeletal muscle;
RX MEDLINE=93036581; PubMed=1329581;
RA Harbitz I., Kristensen T., Bosnes M., Kran S., Davies W.;
RT "DNA sequence of the skeletal muscle calcium release channel cDNA and
RT verification of the Arg615-->Cys615 mutation, associated with porcine
RT malignant hyperthermia, in Norwegian Landrace pigs.";
RL Anim. Genet. 23:395-402(1992).
RN [2]
RP SEQUENCE OF 1129-2801 FROM N.A.
RA Brenig B.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1129-2643 FROM N.A.
RC STRAIN=German Landrace; TISSUE=Liver;
RX MEDLINE=94117003; PubMed=8288238;
RA Leeb T., Schmolzi S., Brem G., Brenig B.;
RT "Genomic organization of the porcine skeletal muscle ryanodine
RT receptor (RYR1) gene coding region 4624 to 7929.";
RL Genomics 18:349-354(1993).
RN [4]
RP SEQUENCE OF 4785-5035 FROM N.A.
RX MEDLINE=91056440; PubMed=21744405;
RA Harbitz I., Chowdhary B., Thomsen P.D., Davies W., Kaufman U.,
RA Kran S., Gustavsson I., Christensen K., Hauge J.G.;

```

"Assignment of the porcine calcium release channel gene, a candidate for the malignant hyperthermia locus, to the 6p11-->q21 segment of chromosome 6."

- GENOMICS 8:243-248 (1990).
- FUNCTION: Communication between transverse-tubules and sarcoplasmic reticulum. Contraction of skeletal muscle is triggered by release of calcium ions from SR following depolarization of T-tubules.
- SUBUNIT: Homotrimer (Potential).
- MISCELLANEOUS: The calcium release channel is modulated by calcium ions, magnesium ions, ATP and calmodulin.
- MISCELLANEOUS: The calcium release channel activity resides in the C-terminal region while the remaining part of the protein constitutes the 'foot' structure spanning the junctional gap between the SR and the T-tubule. It is possible that the foot structure interacts with the cytoplasmic region of the dihydropyridine receptor.
- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-release channel in junctional SR and modulates its activity.
- SIMILARITY: Belongs to the ryanodine receptor family.
- SIMILARITY: Contains 5 MIR domains.
- SIMILARITY: Contains 3 SPRY domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X62880; CAA44674.1; ALT_SEQ.

EMBL; X68247; CAA48318.1; .

EMBL; X63465; CAA49225.1; .

EMBL; M32501; AAA31022.1; .

InterPro: IPR000699; Ca-rel_channel.

InterPro: IPR001682; Ca/Na_pore.

InterPro: IPR002048; EF-hand.

InterPro: IPR003608; Ion_trans.

InterPro: IPR001215; Ryanodn_receptor.

InterPro: IPR003032; RYR.

InterPro: IPR003877; SPRY_receptor.

Pfam: PF00520; Ion_trans; 1.

Pfam: PF02815; MIR; 4.

Pfam: PF01365; RYDR_ITPR; 2.

Pfam: PF02026; RYR; 4.

Pfam: PF00622; SPRY; 3.

PRINTS: PR00795; RYANDINER.

SMART: SMO0472; MIR; 4.

SMART: SMO0449; SPRY; 3.

PROSITE: PS00919; MIR; 5.

Receptor; Transmembrane; Ionic channel; Calcium channel; Repeat; Phosphorylation; Glycoprotein.

DOMAIN 1 3124

TRANSSEM 3125 3145

TRANSSEM 3189 3207

TRANSSEM 3980 3999

TRANSSEM 4018 4036

TRANSSEM 4274 4297

TRANSSEM 4339 4359

TRANSSEM 4557 4578

TRANSSEM 4646 4669

TRANSSEM 4787 4807

TRANSSEM 4835 4854

TRANSSEM 4877 4896

TRANSSEM 4921 4935

DOMAIN 98 153

DOMAIN 160 205

DOMAIN 211 265

DOMAIN 271 334

DOMAIN 336 393

DOMAIN 659 797

FT DOMAIN 1085 1208

FT DOMAIN 1430 1570

FT DOMAIN 1874 1925

FT DOMAIN 4458 4526

FT DOMAIN 842 2960

FT REPEAT 842 955

FT REPEAT 956 1069

FT REPEAT 1345 1360

FT REPEAT 1373 1388

FT REPEAT 2727 2846

FT REPEAT 2847 2960

FT MOD_RES 2844 2844

FT MOD_RES 3947 3947

FT MOD_RES 4320 4320

FT CARBOHYD 3467 3467

FT CARBOHYD 3475 3475

FT CARBOHYD 3904 3904

FT CARBOHYD 3945 3945

FT CARBOHYD 4144 4144

FT CARBOHYD 4862 4862

FT CONFLICT 2892 2092

SO SEQUENCE 5035 AA; 565317 MW; E00613F2027B94A4 CRC64;

Query Match 4.4%; Score 85.5; DB 1; Length 5035;

Best Local Similarity 22.6%; Pred. No. 1.7e+02;

Matches 50; Conservative 24; Mismatches 74; Indels 73; Gaps 12;

Qy 25 VPGW--SRVSGIPSRHP--VPYKRD--FPKPDYQAKYTCPTG----- 68

Db 1001 VAQGSYSYSAVODIPARRN-PLVPEYRLIDBATKRSNDSLCQAVRTLLGYNTIEPPDQ 1059

Qy 69 -SPIPVWEGDDIEVER-----LQAPWFEFKYDGLGKIMDAIGFSTLTGKNTM 121

Db 1060 PSQVESQSRMDRKAIFAEKSYAVQSGRWYFE-----FEAVTIGE-MRV 1102

Qy 122 EWELEFQIGNCTFPHLRPEMDAPFCWQGAACPFEGIDVHWKENGTLVQATISGMFN 181

Db 1103 GWR-----PELRPDEL-----GDELAYVNGRGQRHGLSGELFG 1140

Qy 182 QMAKWKQD-----NEGYYEYVNVKASPEKAEIWF 214

Db 1141 R--FMQSGDVGMIDLTENTIIIFTLNGEVLMSDGSSETAF 1179

RESULT 12

SSA2 MOUSE STANDARD; PRT; 254 AA.

ID SSA2 MOUSE

AC Q99N99;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE 3-oxo-5-alpha-steroid 4-dehydrogenase 2 (EC 1.3.99.5) (Steroid

DE 5-alpha-reductase 2) (SR type 2) (5 alpha-SR2).

GN SRD5A2 OR 5ART2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

RN [1]

SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=21882004; PubMed=11884637;

RA Takeyama K., Kato S.;

RT "Transcriptional regulation of the mouse steroid 5alpha-reductase

RT type II gene by progesterone in brain.";

RL Nucleic Acids Res. 30:1387-1393 (2002).

CC -!- FUNCTION: Converts testosterone into 5-alpha-dihydrotestosterone

CC and progesterone or corticosterone into their corresponding 5-

CC alpha-3-oxosteroids. It plays a central role in sexual

CC differentiation and androgen physiology (By similarity).

CC -!- CATALYTIC ACTIVITY: A 3-oxo-5-alpha-steroid + acceptor = a 3-oxo-

CC delta(4)-steroid + reduced acceptor.

SEQUENCE FROM N.A. (ISOFORM 1).
 MEDLINE=2037893; PubMed=11919186;
 Beer H.-D., Munding C., Dubois N., Mamie C., Hohl D., Werner S.;
 "The estrogen-responsive B box protein: a novel regulator of
 keratinocyte differentiation.";
 J. Biol. Chem. 277:20740-20749(2002).
 [3]
 SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 TISSUE=Lung, Placenta, and Skin;
 MEDLINE=22388257; PubMed=12477932;
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.L., Wang U., Hsieh F.,
 Datchenko L., Maurina M.B., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Ueffing T.B., Toshiyuki S., Carninci P., Prange C.J.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
 Bosak S.A., McMan P.T., McKernan K.J., Malek J.A., Gutierrez P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalon D.K., Murray D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 Butlerfield Y.S.N., Krzyzinski M.I., Skalska U., Smalins D.E.,
 Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 -!- FUNCTION: May play a role in the regulation of keratinocyte
 differentiation.
 -!- SUBCELLULAR LOCATION: Cytoplasmic.
 -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=2;
 Name=1; Synonyms=Alpha;
 IsoId=O95361-1; Sequence=Displayed;
 Name=2; Synonyms=Beta;
 IsoId=O95361-2; Sequence=VSP_009098;
 -!- TISSUE SPECIFICITY: Highest levels found in testis, ovary, small
 intestine, colon, placenta, heart, skeletal muscle and mammary
 gland. More highly expressed in the fetus than in the
 corresponding adult tissues. Expressed in basal keratinocytes.
 -!- SIMILARITY: Belongs to the TRIM/RBCC family.
 -!- SIMILARITY: Contains 1 B30.2-like domain.
 -!- SIMILARITY: Contains 2 B box-type zinc fingers.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

KW Polymorphism.
 FT ZN FING 72 122 B BOX-TYPE 1.
 FT ZN FING 126 165 B BOX-TYPE 2.
 FT DOMAIN 165 203 COILED COIL (POTENTIAL).
 FT DOMAIN 243 274 COILED COIL (POTENTIAL).
 FT DOMAIN 320 340 COILED COIL (POTENTIAL).
 FT DOMAIN 375 534 B30.2-LIKE.
 FT VARSPPLIC 1 216 Missing (in isoform 2).
 FT VARIANT 121 121 E -> D (in dbSNP:2074890).
 FT CONFLICT 235 235 L -> P (IN REF. 1 AND 2).
 FT CONFLICT 246 246 S -> N (IN REF. 1 AND 2).
 FT CONFLICT 465 465 F -> L (IN REF. 3; AAH01564).
 FT CONFLICT 521 521 K -> N (IN REF. 1 AND 2).
 FT CONFLICT 541 541 N -> S (IN REF. 1 AND 2).
 FT CONFLICT 561 561 V -> S (IN REF. 3; AAH01564/AAH15674).
 SQ SEQUENCE 564 AA; 63997 MW; A676070D237D82F8 CRC64;
 Query Match 4.3%; Score 83.5; DB 1; Length 564;
 Best local similarity 22.1%; Pred. No. 16;
 Matches 58; Conservative 31; Mismatches 81; Indels 93; Gaps 16;
 QY 30 SRVSGIPSRHMPVVPYKRPFRKDP-----YCOAKY--TFCPGSPIPVMEGDDIEV 82
 DB 345 TVSAVAVQKRY-----TSKPEPSRSEQFLQAYAYITDP-----DTAHKY 385
 QY 83 PRLQ-----APVPEKYGGLLHKMKMEALIGRSLTGKN-YTMEWY--EFLQIG 130
 DB 386 LRLQENRKYNTTWEHEHPDLPF--RPLH---WRQVLQSOGLYHRYEVELEIFGAG 439
 QY 131 NCTPHLRPEMDAPFWCNGACFEFGIDVHMKENGTLVQVATISGNMFMQAKVXOD 190
 DB 440 TVY-----GLTC--KGIDRKEBN-----SCISGNPFSWLOV--- 471
 QY 191 NETGIYETW-----NYKASPEKGAETWFD-----SYDCKSEVLTFFNKLAE- 232
 DB 472 --NGKEFTAWYSDMETPLKAGFRLGVYIDPPGGLISRYGVETDMLVHFAKCFSEP 529
 QY 233 -FGAFKRIENNYTIFLXSGSP 254
 DB 530 VYAAFWLSKKEKNAIRIVDGESE 552
 RESULT 15
 ID UROM_HUMAN STANDARD; PRT; 640 AA.
 AC P07911;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Uromodulin precursor (Tamm-Horsfall urinary glycoprotein) (THP).
 GN UMOD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87177970; PubMed=3453112;
 RA Pernica D., Kohr W.J., Kuang W.-D., Glaister D., Aggarwal B.B.,
 Chen B.Y., Goedel D.V.;
 RT "Identification of human uromodulin as the Tamm-Horsfall urinary
 glycoprotein.";
 RL Science 236:83-88(1987).
 RN [2]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=87319675; PubMed=3498215;
 RA Hession C., Decker J.M., Sherblom A.P., Kumar S., Yue C.C.,
 Mattaliano R.J., Tizard R., Kawashima E., Schneisener U.,
 Heltky S., Chow E.P., Burne C.A., Shaw A., Muchmore A.V.;
 RA "Uromodulin (Tamm-Horsfall glycoprotein): a renal ligand for
 lymphokines.";

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - protein search, using sw model

on: June 8, 2004, 13:59:08 ; Search time 63.9139 Seconds

(without alignments)
1708.068 Million cell updates/sec

File: US-10-010-050a-2

Sequence: 1927 1 MRGAGAGARASMCWALAL.....IKITVEIPIPIRKITLSGL 346

oring table: BL050W62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

Minimum DB seg length: 0

Minimum DB seg length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP plant:*
10: SP rodent:*
11: SP virus:*
12: SP vertebrate:*
13: SP unclassified:*
14: SP unclassified:*
15: SP virus:*
16: SP bacterioph:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1451	75.3	336	11 Q8R152	Q8R152 mus muscula
2	1429	74.2	322	11 Q8C054	Q8C054 mus muscula
3	1235	72.2	180	5 Q86157	Q86157 dictyosteli
4	202	10.5	378	5 Q86JG6	Q86JG6 dictyosteli
5	101.5	5.3	3118	5 Q17575	Q17575 caenorhabdi
6	101	5.2	633	3 Q8N1H1	Q8N1H1 trichophyto
7	99.5	5.2	514	16 Q8A3V2	Q8A3V2 bacteroides
8	97.5	5.1	682	4 Q8NB06	Q8NB06 homo sapien
9	97.5	5.1	804	5 Q8IAU5	Q8IAU5 plasmodium
10	96.5	5.0	1705	11 Q9ERK5	Q9ERK5 mus muscula
11	96	5.0	508	5 Q8ID54	Q8ID54 plasmodium
12	95.5	5.0	359	10 Q9ZV89	Q9ZV89 arabidopsis
13	95.5	5.0	836	4 Q13785	Q13785 homo sapien
14	95.5	5.0	5322	5 Q9VPL9	Q9VPL9 drosophila
15	94	4.9	362	5 Q9TVM2	Q9TVM2 caenorhabdi
16	94	4.9	1238	16 Q25330	Q25330 helicobacte

17	94	4.9	2313	8 Q8MHW9	Q8MHW9 psilocum nu
18	93	4.8	1349	16 Q8X6C1	Q8X6C1 escherichia
19	92.5	4.8	564	5 Q8WQ12	Q8WQ12 drosophila
20	92.5	4.8	569	5 Q9VEY2	Q9VEY2 drosophila
21	92	4.8	502	16 Q6561	Q6561 aquifex aeo
22	91	4.7	291	5 Q01702	Q01702 caenorhabdi
23	91	4.7	508	10 Q9XBC4	Q9XBC4 arabidopsis
24	91	4.7	607	4 Q9BVZ9	Q9BVZ9 homo sapien
25	91	4.7	742	4 Q94841	Q94841 homo sapien
26	91	4.7	742	4 Q723L6	Q723L6 homo sapien
27	90.5	4.7	850	2 Q8VUF6	Q8VUF6 azarcus ev
28	90.5	4.7	1846	5 Q814Z1	Q814Z1 plasmodium
29	90	4.7	1409	3 Q7Z9F8	Q7Z9F8 coprinus ci
30	89.5	4.6	444	10 Q93ZB3	Q93ZB3 streptococc
31	89.5	4.6	1512	2 Q9WUX5	Q9WUX5 drosophila
32	89	4.6	567	5 Q9GTX3	Q9GTX3 drosophila
33	89	4.6	570	5 Q8M1N6	Q8M1N6 drosophila
34	89	4.6	584	5 Q8MT35	Q8MT35 drosophila
35	89	4.6	613	5 Q9M122	Q9M122 drosophila
36	89	4.6	1353	5 Q9XW65	Q9XW65 caenorhabdi
37	88.5	4.6	465	10 Q9ZS50	Q9ZS50 ipomea bat
38	88.5	4.6	540	2 Q9AIP9	Q9AIP9 methylococc
39	88.5	4.6	752	17 Q9YC61	Q9YC61 aeropyrum p
40	88.5	4.6	1462	4 Q86U07	Q86U07 homo sapien
41	88	4.6	438	3 Q7Z9M6	Q7Z9M6 trichoderma
42	88	4.6	661	12 Q9JG50	Q9JG50 ttv-like mi
43	88	4.6	742	6 Q29397	Q29397 bos taurus
44	88	4.6	2107	12 Q91NQ2	Q91NQ2 tauro syndr
45	88	4.6	2713	5 Q8MMW0	Q8MMW0 dictyosteli

ALIGNMENTS

RESULT 1
Q8R152 PRELIMINARY: PRT; 336 AA.
AC Q8R152. 01-JUN-2002 (TRENBLER. 21, Created)
DT 01-JUN-2002 (TRENBLER. 21, Last sequence update)
DT 01-JUN-2003 (TRENBLER. 24, Last annotation update)
DE Similar to ceroid-lipofuscinosis, neuronal 5 (Fragment).
GN C1N5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; RC025487; AAH25487.1; -;
DR MGD; MGI:2442253; C1N5.
FT NON TER 1
SQ SEQUENCE 336 AA; 38911 MW; D50D74C45483BE1B CRC64;

Query Match 75.3%; Score 1451; DB 11; Length 336;
Best Local Similarity 76.6%; Pred. No. 1.6e-118;
Matches 258; Conservative 31; Mismatches 42; Indels 6; Gaps 2;

QY	12	ASW--CVALALMLAVVGVGSRVSGIPSRHHMPYERKFPDPDPYCOAKYFECTGS	69
QY	4	AHMRPALALMLGLATLGLASPTSG---QRPVYPRFSRFRKTDYCOAKYFECTGS	59
QY	70	PIPYMGDDDIEVFRQLQAPVWEFRYGDLLGHLKIMHDAIGFRSTLTGKNYTMWYELFOL	129
QY	60	PIPYMKNDVIEVLRQAPVWEFRYGDLLGHLKIMHDAVGRSTLTGKNYTMWYELFOL	119
QY	130	GNCTFFPIRPEMDAPFWGNGAACFFPGIDVDVKKNGITLVQVATIGNNFNQAKVYKQ	189
QY	120	GNCTFFPIRPEMDAPFWGNGAACFFPGIDVDVKKNGITLVQVATIGNNFNQAKVYKQ	179
QY	190	DNELTGIVETWNVKASPEKGAETWFDSDSKFVLRTFNKLAEGAFKNIETVYTIIFL	249

2b 180 DNSTGIYIEWTVRADEPGGAQOTWESYDCSNFVLRITKKALEFETREKKELENYTKIFL 235

2y 250 YSGEFPYLLGNSTISVSGPTGNKLTGLAIKRFYFEPKPHLPTEEFLLSLQIPDAVIHKOF 305

2b 240 YSGEFPYLLGNSTISFEPKGNKTLALAIKRFYFEPFPYLYSTKDFLNNFKIPDVIIRKOF 295

2y 310 YLFYFEEYWLFPKKRPFKIKITYEEIPLFRNKTLGL 346

2b 300 YLFYFEEYWLFPKKRPFVKIKITYEEDPLFRKHTFLFDL 336

RESULT 2

ID	Q8C054	PRELIMINARY;	PRT;	322 AA
----	--------	--------------	------	--------

2T 01-MAR-2003 (TEMBLrel. 23, Created)
2T 01-MAR-2003 (TEMBLrel. 23, last sequence update)
2T 01-JUN-2003 (TEMBLrel. 24, last annotation update)
2E Similar to CERIOD-LIPOFUSCINOSIS neuronal protein 5 (Fragment).
3N Cln5.
3S Mus musculus (Mouse).
3C Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
3C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
3X NCBI_TaxId=10090;
3N [1]
3P SEQUENCE FROM N.A.
3C STRAIN=C57BL/6J; TISSUE=Olfactory brain;
3X MEDLINE=22354683; PUBMED=12466851;
3A The FANTOM Consortium.
3A The Riken Genome Exploration Research Group Phase I & II Team;
3T "Analysis of the mouse transcriptome based on functional annotation of
3T 60,770 full-length cDNAs.";
3L Nature 420:563-573(2002).
3R EMBL; AK032293; BAC2797.1; -
3R MGI; MGI:244253; Cln5.
3T NON_TER 1
3Q SEQUENCE 322 AA; 37338 MW; 04FE01B05E0ED8A CRC64;

Query Match	74.2%;	Score 1429;	DB 11;	Length 322;
Post Total similarity	77.0%;	Prod No 1	30.116;	

best local similarity 11.0%; Freq. NO. 1.35-116)
Matches 251; Conservative 31; Mismatches 40; Indels 4; Gaps 1,

[illegible]

RESULT 3

ID	Q86I57	PRELIMINARY;	PRT;	180	AA
AC	Q86I57;				
DT	01-JUN-2003	(Tremblere)	24,	Created)	

DT 01-JUN-2003 (TrEMBLrel_24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel_24, Last annotation update)
DE Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
CC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium
NCBI_Taxid=44689;

RP SEQUENCE FROM N.A.
PC CRYSTALLINITY

RX MEDLINE=22092622; PubMed=12097910;

RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,

RT
RA
tunggal B., Cox E., Quail M.A., Rosenthal A., Noegel A.A.
"Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*":

RL Nature 418:79-85 (2002)

RP SEQUENCE FROM N.A.

RA Baumgart C.;

DB EMBL: AC116305: AA052280.1: -

30 SEQUENCE 180 AA. 30631 MW. 30580CCTAEND1110 CTCC4
KW Hypothetical protein.

Query Match

Best Local Similarity 35.8%; Pred. No. 1.4e-12;
Matches 54; Conservative 25; Mismatches 60; Indels

05 PRDDB-YFOAKY--TECDTGSPIB--VMEGDDTEVERI.OAPVWEERYGDI.I.GHI.KTMHD 1 06

[illegible][illegible]

DB 82 ALGFIYDITIGLNYLEYDAFEEVNGILPNIINVDGDELLMCNAGILCTIVEYINEIYND 141

165 EN---GTLVQVATISGNMENQMAKWKQDN 19

Db 142 KN1YSTASKTYMGTINGTQNLNQYIEMMQYN 17

RESULT 4

086JG6	PRELIMINARY;	PRT;	378	AA
--------	--------------	------	-----	----

DT 01-JUN-2003 (TREMBLrel. 24, Created)

DT 01-JUN-2003 (TREMBLEL, 24, Last annotation update)

Dicystroelium discoideum (S) time mol %).

OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium
 NCBI TaxID=44560.

RN [1] —

```

RC STRAIN=AX4;

```

RA Gloeckner G, Eichinger J, Szafranski K, Pachebat J, Dear P.
MEDLINE=22092622; PubMed=12091910;

RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,

RT "Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*.";

2022-03-03 10:00:00 (2022/03/03 10:00:00) [2] RN

```

RC STRAIN=AX4:
RP SEQUENCE FROM N.A.

```

BA Baumgart C.;
 PI Submitted (MDP-2003) to the EMRI/ConBank/DNRT databases

DR EMBL; AC116982; AA051609.1; --

SEQUENCE 378 AA; 42306 MW; 2CF33D60851FD036 CRC64

Query Match

Best Local Similarity 23.4%; Pred. No. 2,76-09;
Matches 65; Conservative 53; Mismatches 130; Indels 30; Gaps 8


```

Y      228 NKLAEFGAEFFKQNETYTRIFLYSGPTLYGNETSVFGTGNKTLGLAKRFYEPKPHL 287
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b      293 NALAQAN-----FDGGEDYLN-----YRP-NSKNL-----KREYFAPRM 327
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y      288 -PKREFL-ISLQIF-DAVIYKQFPL 311
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b      328 SPKSYIDASVTLQLFYSANIVHDLTYM 354
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 7
b3v2
D      Q8A3V2      PRELIMINARY;      PRT;      514 AA.
C      Q8A3V2;
T      01-JUN-2003 (TrEMBLrel. 24, Created)
T      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
T      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E      Putative beta-xylosidase.
N      B72852.
N      Bacteroides thetaiotaomicron.
K      Bacterioid; Bacteroides; Bacteroides (class); Bacteroidales;
K      Bacteroidaceae; Bacteroides.
X      NCBI_TaxID=818;
X      [1]
X      SEQUENCE FROM N.A.
X      STRAIN=VPI-5482 / ATCC 29148;
X      MEDLINE=22550858; PubMed=12663928;
X      Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
X      Chiang H.C., Hooper L.V., Gordon J.I.,
X      "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
X      Science 299:2074-2076(2003).
X      EMBL; AB016937; AA077958.1;
X      GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
X      GO; GO:0005975; P:carbohydrate metabolism; IEA.
X      InterPro; IPR006710; Glyco_hydro_43.
X      Pfam; PF04616; Glyco_hydro_43; 1.
X      Complete proteome.
X      SEQUENCE 514 AA; 58155 MW; 6F6B8653CC9488 CRC64;

Query Match      5.2%; Score 99.5; DB 16; Length 514;
Best Local Similarity 21.8%; Pred. No. 3.7;
Matches 52; Conservative 30; Mismatches 92; Indels 65; Gaps 9;

Y      41 WPPYKRFDPFRPPDPYCAKYTCFPTGSPDPVM--EGDDIEVFQLQAVWEFKYGD-- 96
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b      294 WPIYGVIDIMNGGEPV--KVTIKPTNGKKVPSFGGSDSPSPSLNIQ--WGFNNPSD 350
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y      97 -----LHGILKINE-----DAIGFRSTLTGNVMEYE-----LPOL 129
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b      351 ADWNLTERKGMILLIKADHLRASHNMLTQKCIYEGYIVTTEMDSWTEGQAGIFCI 410
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y      130 GNCIFPLREMDAPFWCNGGAACFPGI-----DVHMKENGTLVQVATISGMFN 181
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b      411 GN-----LENGIGILKENGKNYVLLNNNGVEKVKVSGKITY 448
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y      182 QMAKWKQDNETGTYETWNVKASP--EKGAETWFDSDYCKFYLRTFNKLAEFGAEFKN 239
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b      449 FRATMANRNTQHLYSTQDKNFTPCGEAASLRFGMKGRVGLYSNTLRDGGNAFFN 507
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 8
b3v2
D      Q8NB36      PRELIMINARY;      PRT;      682 AA.
C      Q8NB36;
T      01-OCT-2002 (TrEMBLrel. 22, Created)
T      01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
T      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E      Hypothetical protein NT2RP2002760.
N      Homo sapiens (human).
N      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
N      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
X      NCBI_TaxID=9606;
X      [1]
X      SEQUENCE FROM N.A.

```

```

RA      Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,
RA      Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,
RA      Nagahari K., Sugano S., Isogai T.;
RA      "HRT human cDNA sequencing project.";
RA      Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
C      EMBL; AK075480; BAC11645.1; -.
DR      GO; GO:0016021; C:integral to membrane; IEA.
DR      GO; GO:0005215; F:transporter activity; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR007114; MFS.
DR      InterPro; IPR005828; Sub-transporter.
DR      InterPro; IPR005829; Sug transporter.
DR      InterPro; IPR005988; SV2.
DR      Pfam; PF00083; sugar tr; 1.
DR      TIGRFAMs; TIGR01289; synapL_SV2; 1.
DR      PROSITE; PS00850; MFS; 1.
DR      PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW      Hypothetical protein; Transmembrane.
X      SEQUENCE 682 AA; 76660 MW; DIA76C11542CC7F CRC64;

Query Match      5.1%; Score 98; DB 4; Length 682;
Best Local Similarity 22.1%; Pred. No. 7.3;
Matches 62; Conservative 40; Mismatches 104; Indels 74; Gaps 16;

```

```

Y      116 GKNYTMWYELPQLGNCCTFPH-----LRPMDAP-FWCNQG-----AACFEGIDDVHMK 164
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b      325 GSAYGFHSRIVFL-VCAEPVSFALGULTQPEPSRFFLENGKHDEAMVVLQVHTNMR 383
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y      165 ENGTLVQVATISG-NMFMQMAKWKQDNETGTYETWNVKASPEKGAETWFDSDYCKFV 223
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b      384 AGSHGERVPSVNIHITIHQEGELIEQSDPTGYQRMGVRL-SLGGQVWGNFLSC----- 438
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y      224 LRTFNKLAEFGAEFFKNIETN-----YTRIFLYSGERTYL-----GNISYVF-- 264
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b      439 -----FGEYRRTILMMGWFTMSFSYGGTLTWPMDIRHQAVDVASRTVFPFG 489
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y      265 -----GPTGNKTLGLAKRFYEPKFP-----HLPTEFLSLIQ--IFDAVIYKQFY-- 310
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b      490 ERVEHTFRTFLNQHRCGQYFNDFGLRLKASFDSDLEBECYFEDVTSSNAFFRNC 549
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y      311 -----LFYN--FEYWFPMKRPFIKITY-----EELPDI 338
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b      550 TEINVFYNTDLEFYEFVNSR--LINSFLHNKEDCELDV 587
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 9
Q8IAU5
ID      Q8IAU5      PRELIMINARY;      PRT;      804 AA.
AC      Q8IAU5;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE      Cullin-1-like protein, putative.
DN      PF08_0094.
OS      Plasmodium falciparum (isolate 3D7).
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX      NCBI_TaxID=36329;
X      [1]
X      SEQUENCE FROM N.A.
X      Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
X      Quail M., Barrrell B.;
X      Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AL844507; CAD51265.1; -.
DR      GO; GO:0007049; P:cell cycle; IEA.
DR      InterPro; IPR001373; Cullin.
DR      Pfam; PF00888; Cullin; 1.
DR      SMART; SM00182; CULLIN; 1.
DR      PROSITE; PS50069; CULLIN_2; 1.
X      SEQUENCE 804 AA; 97162 MW; 6B6E916BA99EAD2C CRC64;

Query Match      5.1%; Score 97.5; DB 5; Length 804;
Best Local Similarity 19.8%; Pred. No. 10;

```



```

92V89
ID Q92V89 PRELIMINARY; PRT; 359 AA.
AC Q92V89;
GT 01-MAY-1999 (TREMBlrel. 10, Created)
GT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
GT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE FPK20.25.
AN FPK20.25.
DS Arabidopsis thaliana (Mouse-ear cress).
DC Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
DC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
DC eurosid1; Brassicales; Brassicaceae; Arabidopsis.
NC NCB1_Taxid=3702;
XN [1]
XP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Vysotskaya V.S., Schwartz J.R., Toriumi M., Yu G., Li J., Liu S.,
RA Kremenetskaya I., Luos J., Araujo R., Buehler E., Conway A.B.,
RA Dewar K., Feng J., Kim C., Li Y., Shinn P., Davis R.W., Ecker J.R.,
RA Federici N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC FPK20 sequence.";
RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005679; AAC83039.1; -.
DR PIR; A96816; A96816.
DR InterPro; IPR004253; DUF231.
DR Pfam; PF03005; DUF231; 1.
SQ SEQUENCE 359 AA; 40568 MW; A9887DD29786278B CRC64;

Query Match 5.0%; Score 95.5; DB 10; Length 359;
Best Local Similarity 21.5%; Pred. No. 5.2;
Matches 53; Conservative 39; Mismatches 97; Indels 57; Gaps 12;

2Y YTPPTGSPPIPVMEGGDDIEVFRLOAPVWEKYGDLGLKIMDAIGFRSTLTGKNYTM 121
80 YRWPPTGCDIPREFGRDFLRFKKTLF--VGDLSLN-----N 116
122 EWEYLFQNGCTPEHLRPEMDAPFCNOGACPF--EGIDVHKENGTLVQVAT--- 174
117 MWVSL-----SCMLHAAPVNAKYTFQLNKGLSTFIPRYGI-SVNFLENGFLVLDVSDKTR 171
2Y 175 ---ISGMFQMAKWKVQDNDEGTGYETWNVKASPEKGAETWPDSPYSCSFVARTPKKA 231
172 GLIKLDSISRGNQMLGSDVAIFNTFHW---SHTGRAKTW-DYFQYGDKIYKNNRME 226
2Y 232 EFGAEF-----KNLETWYTRIFLISGEPYVL-GNETSVFPGTGKTLG--LAIKRFY 280
2b 227 AFKALITWGSKWIDHNDIPSKTRVYGVGSFVHLNGE--WGRPKTCLGETVPVQGPS 283
2Y 281 YPEKPH 286
Db 284 YPGRPN 289

RESULT 13
ID 213785 PRELIMINARY; PRT; 836 AA.
AC Q13785;
GT 01-NOV-1996 (TREMBlrel. 01, Created)
GT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
GT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE APOB protein (Fragment).
AN APOB.
DS Homo sapiens (Human).
DC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NC NCB1_Taxid=9606;
XN [1]
XP SEQUENCE FROM N.A.

```

```

RX MEDLINE=86042646; PubMed=2932736;
RA Wei C.F., Chen S.H., Yang C.Y., Marcel Y.L., Milne R.W., Li W.H.,
RA Sparrow J.T., Gotto A.M. Jr., Chan L.;
RT "Molecular cloning and expression of partial cDNAs and deduced amino
RT acid sequence of a carboxyl-terminal fragment of human apolipoprotein
RT B-100."
RL Proc. Natl. Acad. Sci. U.S.A. 82:7265-7269 (1985).
DR EMBL; M12413; AAA51742.1; -.
DR PIR; A27850; LPHUB.
DR GO; GO:0016020; C-membrane; IEA.
DR GO; GO:0004872; F-transporter activity; IEA.
DR GO; GO:0005215; F-transporter activity; IEA.
DR GO; GO:0006810; P-transport; IEA.
DR InterPro; IPR000531; Trans boxC.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
FT NON_TER 1
SQ SEQUENCE 836 AA; 96317 MW; 5298C27BE21AB140 CRC64;

Query Match 5.0%; Score 95.5; DB 4; Length 836;
Best Local Similarity 20.5%; Pred. No. 16;
Matches 61; Conservative 39; Mismatches 125; Indels 73; Gaps 12;

QY 67 TGSPIPV-----MEGDDIEVFRLOAPVWEKYGDLG--HLKIMDAIGFRSTLTG 116
272 TSAPSPAVGTVMGDMDEDDDFS-----KWNRYSPSSPDKRTIFKTELRVRESDEE 324
QY 117 KNVTMEWYELFQNGCTPEHLRPEMDAPFCNOGACPFEGIDVHKENG--TLVQVAT 175
325 TQIKVMEEBASGLIT--SLKDNV-----KATGVLYDVYKYMHEHGLTLREVSRSK 376
QY 176 SGNMFQMAKWKVQDNDEGTGYETWNVKASPEKGAETWPDSPYSCSFVARTPKKA 235
377 LRRLQDHAEMVYQ-----GAIREDIDDERFQKASGTTGYQ----- 415
QY 236 EFNITNTN-YTRIFLISGEPYVLGNETSVFPGTGKTLGAIK-----RF 279
416 EMDKAKQNIYQELTQEGASFGKDNVEDGLRVVTOEFHMYKHLIDSLDFLNPFRF 475
QY 280 YPEKPHLPTKEFLSLGLQIFDAVIVHKQFY-----LFYFWEYFLPMKFPPIK 328
476 QPFGKRGITREBLCTMF--IREVGTVLSQVYSVNHGSRILSFYFDIVITLLELRK 532

RESULT 14
Q92VPL9 PRELIMINARY; PRT; 5322 AA.
AC Q92VPL9; Q9N164;
GT 01-MAY-2000 (TREMBlrel. 13, Created)
GT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
GT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE C3696 protein (KISMET-L long isoform).
GN KIS OR CG3660 OR CG3696 OR CG18926.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCB1_Taxid=7227;
XN [1]
XP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Man R.H., Doyle C., Baxter B.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Adair J.F., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brostier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

```

de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Dudin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaser K., Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwan C., Jalali M., Kaush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris D.M., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Splitter E., Spradling A.C., Stapleton M., Strong R., Sun E., Svitskas R., Tecor C., Turner R., Venter F., Wang A.H., Wang X., Wang Z.-Y., Massart D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; "The genome sequence of *Drosophila melanogaster*." Science 287:2185-2195 (2000).

[2] SEQUENCE FROM N.A.
Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y., Barton J., An H., Baldwin D., Banzon J., Beeson K.Y., Buesam D.A., Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M., Dodson K., Dorsett V., Doup L.E., Doyle C., Drenek D., Farfan D., Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A., Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J., Ibeagwan C., Jalali M., Kruse D., Li P., Matel B., Moshrefi A., McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J., Pacle J., Paragas V., Park S., Patel S., Pfeiffer B., Phouaneavong S., Pittman G.S., Puri V., Richards S., Scheeler F., Stapleton M., Strong R., Svitskas R., Tecor C., Tyler D., Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.; "Sequencing of *Drosophila melanogaster* genome." Science 287:2185-2195 (2000).

[3] SEQUENCE FROM N.A.
Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K., Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D., Tupy J.L., Bergman C., Bertan B., Carlson J.W., Celinker S.E., Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N., Krommiller B., Marshall B., Millburn G., Richter J., Russo S., Seattle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.I., Lewis S.E.; "Annotation of *Drosophila melanogaster* genome." Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[4] SEQUENCE FROM N.A.
Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.; Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[5] SEQUENCE FROM N.A.
Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

[6] SEQUENCE FROM N.A.
Therrien M., Morrison D.K., Wong A.M., Rubin G.M.; "A Genetic Screen for Modifiers of a KSR-Dependent Rough Eye Phenotype in *Drosophila*." Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
EMBL: A0005590; AAF51527.3; -
EMBL: AF215703; AAF43004.1; -
HSSP: P23197; IAP0.
FlyBase: FBgn001309; kis.
GO: GO:000785; C:chromatin; IEA.

DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO: GO:0003682; F:chromatin binding; IEA.
DR GO: GO:0016787; F:hydrolyase activity; IEA.
DR GO: GO:0006333; P:chromatin assembly/disassembly; IEA.
DR InterPro: IPR000953; Chromo.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR Pfam: PF00385; chromo; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00176; SNF2_N; 1.
DR SMART: SM00298; CHROMO; 2.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC_C; 1.
DR PROSITE: PS50013; CHROMO_2; 1.
DR ATP-binding; Helicase; Hydrolase.
SQ SEQUENCE 5322 AA; 573615 MW; C9608375FA71C211 CRC64;

Query Match 5.0%; Score 95.5; DB 5; Length 5322;
Best Local Similarity 20.1%; Pred. No. 1.8e+02;

Matches 71; Conservative 49; Mismatches 139; Indels 95; Gaps 17;

31 RVSGIPSRHMPVYKPRPKDPYCAKYTCPTGSPFVWBGDDIEVFRLOAPVW 90
1993 RFNKIPQSEW-----KSKKPHPELMKLEKT-----PVYKGNLSRYPQLEGNW 2039

QY 91 EFKYEDDLGLKIMDAIGFSTLTGKNTYMEWYELFOLGNTFPHLPKMAPWNCOG 150
DB 2040 -LKFWYNTNHCILADEMGLGKTIOSLTFVSHVYVGIRG--FPLVIAFLSTIPWQRB- 2095

QY 151 AACFEGGIDDVHMKNGTLVQATISGMFMNMAKWNQDNETGYIYETMNVKASPE--- 207
DB 2096 ---PEG---WTDNVAVVYHGSVT-----SKQMTQDVE--YYKTESKVLKEPIK 2137

QY 208 -KGAETWDSYDCSKFVLRFTEN-----KLAEGAFENIE----- 241
DB 2138 FNVLLTFEMVLYTMDLKAFWRLCVIDEARLRKNRCKLIE-GLRLQNEHRVLLSGT 2196

QY 242 --TNYTRIF--LYSGEPFYGNETSVPFGPTGNKTLGLAIKFPYEPFHL----- 287
DB 2197 PLQNNISEFLSNLFLSEQSFSSQEFMSFSLTTEERVNQLQALIKPMVRLKQDVE 2256

QY 288 ----PTKEFLSLDQIPDAVIHKQFYLFPNRYWFLPKKPEFIK--ITYEELP 335
DB 2257 KSLAKPEETILRV---ELTNQKKRY-----RGLPEQNFSLKKGTTSANIP 2300

RESULT 15
QY 09TYM2 PRELIMINARY; PRT; 362 AA.
AC 09TYM2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Y25C1A.7a protein.
GN Y25C1A.7 OR Y25C1A.7A.

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;

SEQUENCE FROM N.A.
RC STRAIN=Bristol N2.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., Gardner A., Green P., Green T., Hillier L., Jier M., Johnston L., Jones M., Kerhaw J., Kirsten J., Laister N., Lattelle P., Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M., Parsons J., Percy C., McKurray L., Koopra A., Saunders D., Showkneen R.,

```

1A Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
2A Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
3A Watson A., Weinstock L., Wilkinson-Spratt J., Wooldman P.:
4A "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
5A elegans.";
6A Nature 368:32-38(1994).
7A
8A [2]
9A SEQUENCE FROM N.A.
10A STRAIN=Bristol N2;
11A Kalicki J., Smith A., Gibson A.:
12A "The sequence of C. elegans cosmid Y25C1A.";
13A Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
14A [3]
15A SEQUENCE FROM N.A.
16A STRAIN=Bristol N2;
17A Waterston R.;
18A Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
19A EMBL; AF125459; AAD12838.1; -.
20A PIR; T33904; T33904.
21A WormPep; Y25C1A.7a; CE21474.
22A InterPro; IPR006977; DUF649.
23A Pfam; PF04893; Yip1; 1.
24A SEQUENCE 362 AA; 40613 MW; 42DB45A7F72F3B89 CRC64;
25A
26A
27A
28A
29A
30A
31A
32A
33A
34A
35A
36A
37A
38A
39A
40A
41A
42A
43A
44A
45A
46A
47A
48A
49A
50A
51A
52A
53A
54A
55A
56A
57A
58A
59A
60A
61A
62A
63A
64A
65A
66A
67A
68A
69A
70A
71A
72A
73A
74A
75A
76A
77A
78A
79A
80A
81A
82A
83A
84A
85A
86A
87A
88A
89A
90A
91A
92A
93A
94A
95A
96A
97A
98A
99A
100A
101A
102A
103A
104A
105A
106A
107A
108A
109A
110A
111A
112A
113A
114A
115A
116A
117A
118A
119A
120A
121A
122A
123A
124A
125A
126A
127A
128A
129A
130A
131A
132A
133A
134A
135A
136A
137A
138A
139A
140A
141A
142A
143A
144A
145A
146A
147A
148A
149A
150A
151A
152A
153A
154A
155A
156A
157A
158A
159A
160A
161A
162A
163A
164A
165A
166A
167A
168A
169A
170A
171A
172A
173A
174A
175A
176A
177A
178A
179A
180A
181A
182A
183A
184A
185A
186A
187A
188A
189A
190A
191A
192A
193A
194A
195A
196A
197A
198A
199A
200A
201A
202A
203A
204A
205A
206A
207A
208A
209A
210A
211A
212A
213A
214A
215A
216A
217A
218A
219A
220A
221A
222A
223A
224A
225A
226A
227A
228A
229A
230A
231A
232A
233A
234A
235A
236A
237A
238A
239A
240A
241A
242A
243A
244A
245A
246A
247A
248A
249A
250A
251A
252A
253A
254A
255A
256A
257A
258A
259A
260A
261A
262A
263A
264A
265A
266A
267A
268A
269A
270A
271A
272A
273A
274A
275A
276A
277A
278A
279A
280A
281A
282A
283A
284A
285A
286A
287A
288A
289A
290A
291A
292A
293A
294A
295A
296A
297A
298A
299A
300A
301A
302A
303A
304A
305A
306A
307A
308A
309A
310A
311A
312A
313A
314A
315A
316A
317A
318A
319A
320A
321A
322A
323A
324A
325A
326A
327A
328A
329A
330A
331A
332A
333A
334A
335A
336A
337A
338A
339A
340A
341A
342A
343A
344A
345A
346A
347A
348A
349A
350A
351A
352A
353A
354A
355A
356A
357A
358A
359A
360A
361A
362A
363A
364A
365A
366A
367A
368A
369A
370A
371A
372A
373A
374A
375A
376A
377A
378A
379A
380A
381A
382A
383A
384A
385A
386A
387A
388A
389A
390A
391A
392A
393A
394A
395A
396A
397A
398A
399A
400A
401A
402A
403A
404A
405A
406A
407A
408A
409A
410A
411A
412A
413A
414A
415A
416A
417A
418A
419A
420A
421A
422A
423A
424A
425A
426A
427A
428A
429A
430A
431A
432A
433A
434A
435A
436A
437A
438A
439A
440A
441A
442A
443A
444A
445A
446A
447A
448A
449A
450A
451A
452A
453A
454A
455A
456A
457A
458A
459A
460A
461A
462A
463A
464A
465A
466A
467A
468A
469A
470A
471A
472A
473A
474A
475A
476A
477A
478A
479A
480A
481A
482A
483A
484A
485A
486A
487A
488A
489A
490A
491A
492A
493A
494A
495A
496A
497A
498A
499A
500A
501A
502A
503A
504A
505A
506A
507A
508A
509A
510A
511A
512A
513A
514A
515A
516A
517A
518A
519A
520A
521A
522A
523A
524A
525A
526A
527A
528A
529A
530A
531A
532A
533A
534A
535A
536A
537A
538A
539A
540A
541A
542A
543A
544A
545A
546A
547A
548A
549A
550A
551A
552A
553A
554A
555A
556A
557A
558A
559A
560A
561A
562A
563A
564A
565A
566A
567A
568A
569A
570A
571A
572A
573A
574A
575A
576A
577A
578A
579A
580A
581A
582A
583A
584A
585A
586A
587A
588A
589A
590A
591A
592A
593A
594A
595A
596A
597A
598A
599A
600A
601A
602A
603A
604A
605A
606A
607A
608A
609A
610A
611A
612A
613A
614A
615A
616A
617A
618A
619A
620A
621A
622A
623A
624A
625A
626A
627A
628A
629A
630A
631A
632A
633A
634A
635A
636A
637A
638A
639A
640A
641A
642A
643A
644A
645A
646A
647A
648A
649A
650A
651A
652A
653A
654A
655A
656A
657A
658A
659A
660A
661A
662A
663A
664A
665A
666A
667A
668A
669A
670A
671A
672A
673A
674A
675A
676A
677A
678A
679A
680A
681A
682A
683A
684A
685A
686A
687A
688A
689A
690A
691A
692A
693A
694A
695A
696A
697A
698A
699A
700A
701A
702A
703A
704A
705A
706A
707A
708A
709A
710A
711A
712A
713A
714A
715A
716A
717A
718A
719A
720A
721A
722A
723A
724A
725A
726A
727A
728A
729A
730A
731A
732A
733A
734A
735A
736A
737A
738A
739A
740A
741A
742A
743A
744A
745A
746A
747A
748A
749A
750A
751A
752A
753A
754A
755A
756A
757A
758A
759A
760A
761A
762A
763A
764A
765A
766A
767A
768A
769A
770A
771A
772A
773A
774A
775A
776A
777A
778A
779A
780A
781A
782A
783A
784A
785A
786A
787A
788A
789A
790A
791A
792A
793A
794A
795A
796A
797A
798A
799A
800A
801A
802A
803A
804A
805A
806A
807A
808A
809A
810A
811A
812A
813A
814A
815A
816A
817A
818A
819A
820A
821A
822A
823A
824A
825A
826A
827A
828A
829A
830A
831A
832A
833A
834A
835A
836A
837A
838A
839A
840A
841A
842A
843A
844A
845A
846A
847A
848A
849A
850A
851A
852A
853A
854A
855A
856A
857A
858A
859A
860A
861A
862A
863A
864A
865A
866A
867A
868A
869A
870A
871A
872A
873A
874A
875A
876A
877A
878A
879A
880A
881A
882A
883A
884A
885A
886A
887A
888A
889A
890A
891A
892A
893A
894A
895A
896A
897A
898A
899A
900A
901A
902A
903A
904A
905A
906A
907A
908A
909A
910A
911A
912A
913A
914A
915A
916A
917A
918A
919A
920A
921A
922A
923A
924A
925A
926A
927A
928A
929A
930A
931A
932A
933A
934A
935A
936A
937A
938A
939A
940A
941A
942A
943A
944A
945A
946A
947A
948A
949A
950A
951A
952A
953A
954A
955A
956A
957A
958A
959A
960A
961A
962A
963A
964A
965A
966A
967A
968A
969A
970A
971A
972A
973A
974A
975A
976A
977A
978A
979A
980A
981A
982A
983A
984A
985A
986A
987A
988A
989A
990A
991A
992A
993A
994A
995A
996A
997A
998A
999A
1000A

```

Search completed: June 8, 2004, 14:07:04
 Job time : 66.9139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - protein search, using sw model

on: June 8, 2004, 13:55:43 ; Search time 7.42778 Seconds
(without alignments)
1065.101 Million cell updates/sec

File: US-10-010-050a-2_COPY_1_28
154

Project score: 1 MRRGAGARGASWCMTALIMLAVPG 28

Sequence: BLOSUM62

oring table: Gapop 10.0 , Gapext 0.5

Marched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	347	2	AAV41323 Human sec
2	154	100.0	358	4	AAW93870 Human pol
3	147	95.5	346	2	AAW92967 Human zbl
4	60.5	39.3	192	4	AAU29418 Human G P
5	60.5	39.3	192	5	ABG60706 Novel G P
6	59.5	38.6	127	4	AAU14162 Human nov
7	58	37.7	188	5	ABU60973 Lung spec
8	57.5	37.3	19938	6	ABP76682 Streptomy
9	56	36.4	214	5	ABU05547 M. tuberc
10	55	35.7	403	5	ABU65190 Human NOV
11	54	35.1	78	4	ABU16658 Human ner
12	53	34.4	34	2	AAV43189 Thioeredox
13	53	34.4	211	4	ABG01709 Novel hum
14	53	34.4	346	5	ABG96249 Naize per
15	53	34.4	600	4	ABG07764 Novel hum
16	52.5	34.1	199	4	AAU28370 Novel hum
17	52.5	34.1	199	7	AAU09107 Novel pro
18	52.5	34.1	199	7	AAU09108 Novel pro
19	52.5	34.1	267	4	AAW92968 Corynebact
20	52.5	34.1	333	3	AAW91017 C glutami
21	52.5	34.1	341	3	AAW92980 B. lactof
22	52.5	34.1	360	6	ABU22472 Protein e
23	52.5	34.1	442	1	AAW90475 N-termina
24	52.5	34.1	551	2	AAW76503 Cattle MI
25	52.5	34.1	575	1	AAW70195 Sequence

ALIGNMENTS

RESULT 1	AAV41323	standard; protein; 347 AA.
XX	AAV41323	
XX	AAV41323	
AC	AAV41323	
DT	02-DEC-1999	(first entry)
XX	Human secreted protein encoded by gene 16 clone HMZAD77.	
XX	Human, secreted protein; fusion protein; gene therapy; protein therapy;	
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukoemia;	
KW	developmental abnormality; foetal deficiency; blood; allergy; renal;	
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;	
KW	inflammation; ischaemic shock; Alzheimer's disease; osteoclast; thymus;	
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;	
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;	
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.	
XX		
OS	Homo sapiens.	
XX		
PN	WO9947540-A1.	
XX		
PD	23-SEP-1999.	
XX		
PF	18-MAR-1999;	99WO-US005804.
XX		
PR	19-MAR-1998;	98US-0078563P.
PR	19-MAR-1998;	98US-0078566P.
PR	19-MAR-1998;	98US-0078573P.
PR	19-MAR-1998;	98US-0078574P.
PR	19-MAR-1998;	98US-0078576P.
PR	19-MAR-1998;	98US-0078577P.
PR	19-MAR-1998;	98US-0078578P.
PR	19-MAR-1998;	98US-0078579P.
PR	19-MAR-1998;	98US-0078581P.
PR	01-APR-1998;	98US-0080312P.
PR	01-APR-1998;	98US-0080313P.
PR	01-APR-1998;	98US-0080314P.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;	
PI	Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Latleur DW, Olsen HS;	
PI	Shi Y, Moore PA,	
XX	WPI, 1999-562050/47.	
DR	N-PSDB; AA24826.	

New isolated human genes, useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders.
 Claim 11, Page 367-368, 484pp; English.
 This sequence represents a secreted human protein encoded by the gene clone detailed in the descriptor line. The gene can be used to generate fusion proteins by linking to the gene to a human immunoglobulin Fc portion (e.g. AA224802) for increasing the stability of the fused protein as compared to the human protein only. The invention relates to 95 novel genes and their fragments (nucleic acid sequences: AA224811-624907; amino acid sequences AA541308-Y14404) which are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. Also, pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the 95 polynucleotides, based on which tissues they are most highly expressed in (see AA224811 for described uses)
 Sequence 347 AA:
 Query Match 100.0%; Score 154; DB 2; Length 347;
 Best Local Similarity 100.0%; Pred. No.3.1e-13;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0
 1 MRRGAGARGRASMCMALALMLAVVG 28
 1 MRRGAGARGRASMCMALALMLAVVG 28
 RESULT 2
 I.D. AA093870
 I.D. AA093870 strand; protein; 358 AA.
 AA093870;
 06-NOV-2001 (first entry)
 Human polypeptide, SEQ ID NO: 3978.
 Human; full length cDNA; cDNA synthesis; oligo-capping.
 Homo sapiens.
 EP1130094-A2.
 05-SEP-2001.
 07-UTL-2000; 2000EP-00114089.
 08-UTL-1999; 99CP-00194486.
 11-JAN-2000; 2000JP-00118774.
 02-MAY-2000; 2000JP-00183765.
 (HELI-) HELIX RES INST.
 Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y, H, Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H, MFI; 2001-524255/58.
 N-PSDB; AKK94829.
 830 Primers useful for synthesizing full length cDNA clones and their use in genetic manipulation.
 Claim 8; SEQ ID NO 3978; 1380bp + Sequence Listing; English.
 The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful

CC	for clarifying the function of the protein encoded by the cDNA. The full
CC	length clones were obtained by construction of full length enriched cDNA
CC	libraries that were synthesised by the oligo-capping method. The primers
CC	enable the production of the full length cDNA easily without any special
CC	methods. The present sequence is a polypeptide encoded by a full length
CC	human cDNA of the invention. Note: The sequence data for this patent did
CC	not form part of the printed specification, but was obtained in CD-ROM
CC	format directly from EPO
XX	
XX	Sequence 358 AA;
SQ	
Query Match	100.0%; Score 154; DB 4; Length 358;
Best Local Similarity	100.0%; Pred. No. 3 2e-13;
Matches 28; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 MRRGAGARGRASWCWALATLMTAVVVG 28 13 MRRGAGARGRASWCWALATLMTAVVVG 40
Dn	
RESULT 3	
ID	AAM92967 standard; protein; 346 AA.
AAW92967	
AC	AAM92967;
DT	14-MAY-1999 (first entry)
XX	
DE	Human zsig46 protein.
XX	
MW	Secreted protein; zsig46; human; chromosome 13; thyroid disease;
KM	hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;
KW	Hirschsprung's disease; neuronal ceroid-lipofuchnosis; Wilson disease;
RW	Reiger syndrome; immunoassay; detection; anti-idiotypic antibody;
KW	therapy; diagnostic.
XX	
OS	Homo sapiens.
PN	WO9905275-A1.
FD	04-FEB-1999.
PF	24-JUL-1998; 98WO-US015431.
XX	
PR	24-JUL-1997; 97US-0053613P.
PA	(ZYMO) ZYMOGENETICS INC.
PI	Sheppard PO, Gilbertson DG;
DR	WI: 1999-142930/12.
N-PSDB; AAX02855.	
New secreted polypeptide, zsig46, and its fragments, related fusion	
proteins - used for diagnosis and treatment of thyroid disorders or	
diseases involving genes on chromosome 13.	
Claim 3; Page 90-91; 10tp; English.	
This invention describes the isolation of a novel human secreted protein,	
zsig46 encoded by a gene on chromosome 13 which is mainly expressed in	
the thyroid. This product can be used to study secretion of proteins from	
cells and also to treat or prevent deficient expression of zsig46, which	
may be associated with thyroid diseases (e.g. hypothyroidism, Graves'	
disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that	
involve genes in the same region of chromosome 13 (e.g. Hirschsprung's	
disease, neuronal ceroid-lipofuchnosis, Wilson disease and Reiger	
syndrome). Antibodies and other binding proteins, are used as immunoassay	
reagents to detect zsig46 or cells expressing it, e.g. for assessing	
thyroid function to produce anti-idiotypic antibodies, for affinity	
purification of zsig46, to screen expression libraries, to neutralise	
zsig46 activity, and to deliver toxins, radioisotopes etc. for	
therapeutic or diagnostic purposes. Agonists of the product can be used	

to promote growth, differentiation and proliferation of specific cell types, e.g. for treating (extra)thyroid diseases or as additive to cell cultures

Sequence 346 AA;

Query Match 95.5%; Score 147; DB 2; Length 346;

Best Local Similarity 96.4%; Pred. No. 2.9e-12;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MRRGAGARGRSMCWALALMLAVPG 28

1 MGRGAGARGRSMCWALALMLAVPG 28

AAU29418 standard; protein; 192 AA.

AAU29418;

18-DEC-2001 (first entry)

Human G protein-coupled receptor (GPCR) polypeptide #39.

Human; G protein-coupled receptor; GPCR; mental disorder; schizophrenia; neurological disorder; metabolic disorder; cancer; rheumatoid arthritis; thyroid disorder; neurodegenerative disorder; cardiovascular disorder; renal failure; autoimmune disorder; hyperproliferative disorder; HIV; human immunodeficiency virus; viral infection; neuroprotective; immunostimulant; neuroleptic; nootropic; tranquiliser; antidepressant; anorectic; gene therapy.

Homo sapiens.

WO200168858-A2.

20-SEP-2001.

16-MAR-2001; 2001WO-US008456.

16-MAR-2000; 2000US-0187783P.

16-MAR-2000; 2000US-0189907P.

16-MAR-2000; 2000US-0189917P.

16-MAR-2000; 2000US-0189918P.

16-MAR-2000; 2000US-0189960P.

29-MAR-2000; 2000US-0192155P.

29-MAR-2000; 2000US-0192234P.

29-MAR-2000; 2000US-0192830P.

29-MAR-2000; 2000US-0192830P.

29-MAR-2000; 2000US-0192833P.

29-MAR-2000; 2000US-0192933P.

29-MAR-2000; 2000US-0192945P.

(PHAA) PHARMACIA & UPJOHN CO.

Vogeli G;

WPI; 2001-607458/69.

N-PSDB; AAS46857.

Nucleic acid encoding G-protein coupled receptors, useful for the prevention, diagnosis and treatment of mental disorders.

Claim 31; Page 73; 274pp; English.

Sequences AAU29380-AAU29509 represent human G protein-coupled receptor (GPCR) polypeptides of the invention. The proteins and the DNA sequences encoding them can be used to identify compounds which bind to GPCR

polypeptides and in screening for compounds that modulate GPCR activity.

By screening a human subject for the presence of mutations in GPCR DNA, a GPCR-related disorder or a genetic predisposition can be diagnosed. The

sequences can also be used for treatment and prevention of mental

disorders such as schizophrenia, neurological disorders such as manic depression, metabolic disorders such as obesity, cancer, rheumatoid arthritis, thyroid disorders such as myxoedema, neurodegenerative disorders such as Parkinson's disease, cardiovascular disorders such as atherosclerosis, renal failure, autoimmune disorders, hyperproliferative disorders such as psoriasis and viral infections such as those caused by HIV

Sequence 192 AA;

Query Match

Best Local Similarity 39.3%; Score 60.5; DB 4; Length 192;

Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

4 GAGARGRSMCW-----ALALLW 22

35 GKAAVGRPCWPCQPALVSTALVW 62

ABG60706

ABG60706;

13-AUG-2002 (first entry)

Novel G protein coupled receptor (nGPCR-x) #39.

G protein coupled receptor; nGPCR-x; immune response; thyroid disorder; mental disorder; chylotoxocosis; myxoedema; inflammatory condition; Crohn's disease; cell differentiation; homeostasis; rheumatoid arthritis; renal failure; autoimmune disorder; movement disorder; CNS disorder; viral infection; human immunodeficiency virus; HIV; metabolic disorder; cardiovascular disorder; diabetes; obesity; anorexia; cardiomyopathy; poriferative disease; cancer; psoriasis; lung cancer; hormonal disorder; sexual dysfunction.

Homo sapiens.

US2002058306-A1.

16-MAY-2002.

16-MAR-2001; 2001US-00811284.

16-MAR-2000; 2000US-0189783P.

16-MAR-2000; 2000US-0189907P.

16-MAR-2000; 2000US-0189917P.

16-MAR-2000; 2000US-0189918P.

16-MAR-2000; 2000US-0189960P.

24-MAR-2000; 2000US-0192155P.

27-MAR-2000; 2000US-0192234P.

29-MAR-2000; 2000US-0192830P.

29-MAR-2000; 2000US-0192916P.

29-MAR-2000; 2000US-0192923P.

29-MAR-2000; 2000US-0192933P.

29-MAR-2000; 2000US-0192935P.

29-MAR-2000; 2000US-0192945P.

(VOGE/) VOGELI G.

Vogeli G;

WPI, 2002-434856/46.

N-PSDB; ABR61635.

New isolated nucleic acid encoding a G protein coupled receptor for

producing the receptor which can induce an immune response in a mammal.

Claim 27; Page 44; 216pp; English.

The invention describes an isolated nucleic acid (1) comprising a

sequence encoding a portion of a G protein coupled receptor (ngPCR-x).
(1) is used to produce a recombinant ngPCR-x polypeptide. A polypeptide encoded by (1) is used to induce an immune response in a mammal. ngPCR-x is used to identify a compound that binds to it and/or modulates its activity. (1) is used to identify animal homologues of ngPCR-x. (1) can be used to diagnose a human subject as having a brain or genetic predisposition disorder, such as a mental disorder. (1) is used to screen for an ngPCR-x related disorder including thyroid disorders (e.g. thyrotoxicosis, myxoedema), renal failure, inflammatory conditions (e.g. Crohn's disease), diseases related to cell differentiation and homeostasis, rheumatoid arthritis, autoimmune disorders, movement disorders, CNS disorders, viral infections (e.g. diabetes, obesity, virus), metabolic and cardiovascular disorders (e.g. diabetes, obesity, anorexia, cardiomyopathies), poriferative diseases and cancers (e.g. psoriasis, lung cancer), hormonal disorders, sexual dysfunction and hereditary mental disorders in a human patient. A host cell comprising (1) is used to screen for a modulator of ngPCR-x activity. ngPCR-x is used to identify compounds that can treat mental disorders. The polypeptide encoded by (1) is used to purify a G protein from a sample. This is the amino acid sequence of a novel G protein coupled receptor (ngPCR-x) protein described in the invention

Sequence 192 AA:

Query Match 39.3%; Score 60.5; DB 5; Length 192;
Best Local Similarity 39.3%; Pred. No. 1.8;
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

Y 4 GAGAGRGASWCW-----ALALILW 22
35 GKAAVGRPCWPCQPALVSIILAW 62

RESULT 6
AAU14162 standard; protein; 127 AA.

AAU14162;

24-OCT-2001 (first entry)

Human novel protein #33.

Human, novel protein; Anti-anemic; osteopathic; anti-inflammatory;
immunomodulatory; cytosolic; neuroprotective; valnerary; nootropic;
anticonvulsant; antiarrhythmic; cerebroprotective; antitumoral; antiviral;
antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
tissue regeneration; immune disorder.

Homo sapiens.

WO200155437-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US002623.

25-JAN-2000; 2000US-00491404.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-451939/48.

N-PSDB; AAS22467.

Isolated polypeptides useful for treating anti-inflammatory diseases,
nervous system disorders, and for regenerating bone and cartilage.

Example 4; Page 543-544; 894pp; English.

The invention relates to polynucleotides encoding novel human proteins or their active domains. The polypeptides, polynucleotides and antibodies raised against the polypeptides are used in a method of treatment of a mammal and prevention of disorders caused by the aberrant protein expression or activity. The polypeptides can be used as molecular weight markers, food supplements, and in antibody production. The polypeptides are used to identify compounds which bind to the polypeptides. Polynucleotides of the invention are used as probes and primers, for sequencing, for chromosome or gene mapping, in the production of recombinant proteins, and in generating anti-sense DNA or RNA and in gene therapy. Polypeptides of the invention can be used to target drugs to a tumour, in assays to determine biological activity, to raise antibodies/elicite an immune response, to determine quantitative protein levels, as tissue markers, and to isolate receptors or ligands. Polypeptides of the invention may also be useful in treating platelet disorders, stem cell disorders, regenerating bone, cartilage, tendon, ligament and/or nerve tissue, wound healing, treating burns, promoting the proliferation, differentiation and survival of stem cells, as a contraceptive, treating osteoporosis and osteoarthritis, anaemia, Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral sclerosis, stroke, immune deficiencies resulting from bacterial, viral or fungal infection or from autoimmunity, cancer, allergy, asthma, graft-versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory diseases, nervous system disorders, and infection. The present sequence represents a protein of the invention

Sequence 127 AA:

Query Match 38.6%; Score 59.5; DB 4; Length 127;
Best Local Similarity 40.0%; Pred. No. 1.6;
Matches 14; Conservative 3; Mismatches 9; Indels 9; Gaps 2;

QY 3 RGAGARGR-----ASWCALALILW---LAVVPG 28
25 RRGRTWRVMTKIAQMLGAILIGSTVALITG 59

RESULT 7
ABU60973 standard; protein; 188 AA.

ABU60973;

08-MAY-2003 (first entry)

Lung specific protein (LSP) #76.

Human; gene therapy; vaccine; lung specific antigen; cancer diagnosis;
cancer monitoring; cancer staging; cancer imaging; lung cancer;
non-cancerous diseases of the lung; transgenic animal.

Homo sapiens.

WO200268633-A2.

06-SEP-2002.

21-NOV-2001; 2001WO-US043612.

22-NOV-2000; 2000US-0252500P.

(DIAD-) DIADEXUS INC.

Macina RA, Recipon H, Chen S, Sun Y, Liu C;

WPI; 2002-713376/77.

Isolated human nucleic acid molecule and polypeptide, useful for
identifying, diagnosing, monitoring, staging, imaging and treating lung
cancer and non-cancerous diseases of the lung.

Claim 11; Page 359-360; 389pp; English.

The invention describes an isolated human nucleic acid (I) encoding any of 120 10-1533 residue amino acid sequences (S1), given in the specification, comprising any of 164 179-12421 base pair sequences (S2), given in the specification. The methods and compositions of the present invention are useful for identifying, diagnosing, monitoring, staging, imaging and treating lung cancer and non-cancerous diseases of the lung. They are also used for identifying lung tissue, monitoring and identifying and/or designing antagonists of the polypeptide of the invention, gene therapy, production of transgenic animals and production of engineered lung tissue for treatment and research. This is the amino acid sequence of a lung specific nucleic acid

Q Sequence 188 AA;

Query Match 37.7%; Score 58; DB 5; Length 188;

Best Local Similarity 38.7%; Pred. No. 3.9;

Matches 12; Conservative 1; Mismatches 12; Indels 6; Gaps 1;

Y 4 GAGARGRASWCM-----ALALMLAVPG 28
19 GGGGAGAGSWWGMGSGAGALMVAVVG 49

RESULT 8
BP76682

D ABP76682 standard; protein; 19938 AA.

C ABP76682;

T 26-FEB-2003 (first entry)

E Streptomyces viridochromogenes Avi gene cluster polypeptide frame 6.

W Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection; medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.

S Streptomyces viridochromogenes.

N WO200268436-A1.

D 06-SEP-2002.

F 24-AUG-2001; 2001WO-EP009815.

R 25-FEB-2001; 2001DE-01009166.

A (COMB-) COMBINATURE BIOPHARM AG.

I Weitzner G, Muehlenweg A, Trefzer A, Bechtold A;

R MPI; 2003-018650/01.

R N-PSDB; AB237516.

T New avilamycin derivatives, useful for treatment of infections, and nucleic acid encoding avilamycin synthesis enzymes.

S Example 1; Page 68-301; 319pp; German.

The invention relates to avilamycin derivatives (I) with antibacterial, virucide, protozoacide and fungicide activity. (I) are useful for treatment of infections (bacterial, viral, protozoal or fungal), in human or veterinary medicine, particularly where caused by Staphylococcus aureus. (I) are more hydrophilic than known avilamycins. The present sequence is that of an avilamycin synthesis enzyme from the Streptomyces viridochromogenes Avilamycin A biosynthetic gene cluster (AB237515-AB237516)

Q Sequence 19938 AA;

Query Match 37.3%; Score 57.5; DB 6; Length 19938;

Best Local Similarity 41.4%; Pred. No. 5.3e+02;

Matches 12; Conservative 3; Mismatches 11; Indels 3; Gaps 1;

OY 2 RGARASWCMALMLAVP 27
DB 2905 RRAATPRGARGVEMCGPVGWLCAP 2933

RESULT 9
AB05547

ID AB05547 standard; protein; 214 AA.

AC AB05547;

DT 08-APR-2003 (first entry)

DE M. tuberculosis and M. leprae marker protein #198.

XX Mycobacterioses; survival; virulence; protective antigen; vaccine;

KW Mycobacterial disease; tuberculosis; leprosy.

OS Mycobacterium tuberculosis.

PN WO200274903-A2.

PD 26-SEP-2002.

PF 22-FEB-2002; 2002MO-IB001973.

PR 22-FEB-2001; 2001US-0270123P.

PA (INSP) INST PASTEUR.

PI Cole S;

DR MPI; 2002-759885/82.

PT Identifying and selecting genes for survival or virulence of mycobacteria

PT by a comparative genomic analysis of the sequences of Mycobacterium

PS tuberculosis and M. leprae.

PS Claim 17; Page 376-377; 874pp; English.

This invention relates to a novel method for identifying essential genes for survival or virulence of mycobacteria species. The method comprises aligning the genomic sequence of a first mycobacterium species and a genomic sequence of a second mycobacterium species and selecting a CC polynucleotide sequence that is highly conserved in both genomes with no CC counterparts in other bacterial genomic sequences and that corresponds to an essential gene for the survival or virulence of mycobacterium species. CC The method of the invention is useful for detecting M. tuberculosis or M. leprae infection. The method reduces the number of potential new targets and protective antigens for new drugs and vaccine compositions to treat CC and prevent mycobacterial diseases, particularly tuberculosis and CC leprosy. The present sequence represents a marker protein from CC Mycobacterium tuberculosis and Mycobacterium leprae identified using the CC method of the invention

SQ Sequence 214 AA;

Query Match 36.4%; Score 56; DB 5; Length 214;

Best Local Similarity 41.7%; Pred. No. 8.4;

Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

OY 3 RGARASWCMALMLAVP 26
DB 175 RTSATRGYQWRMSIATIMFTV 198

RESULT 10

ID AB065190 standard; protein; 403 AA.

AC AB065190;

XX

20-MAY-2003 (first entry)
Human NOV105a protein.
NOVX; cytosolic; cardiant; antiarteriosclerotic; antiasthmatic; cancer;
hypocensive; cardiomyopathy; bronchial asthma; gene therapy; vaccine;
human.
Homo sapiens.
MO200272757-A2.
19-SEP-2002.
08-MAR-2002; 2002MO-US006908.
08-MAR-2001; 2001US-0274101P.
08-MAR-2001; 2001US-0274194P.
08-MAR-2001; 2001US-0274281P.
08-MAR-2001; 2001US-0274322P.
09-MAR-2001; 2001US-0274849P.
12-MAR-2001; 2001US-0275235P.
13-MAR-2001; 2001US-0275578P.
13-MAR-2001; 2001US-0275579P.
13-MAR-2001; 2001US-0275601P.
14-MAR-2001; 2001US-0276000P.
16-MAR-2001; 2001US-0276776P.
19-MAR-2001; 2001US-0276994P.
20-MAR-2001; 2001US-0277239P.
20-MAR-2001; 2001US-0277327P.
21-MAR-2001; 2001US-0277919P.
22-MAR-2001; 2001US-0278332P.
23-MAR-2001; 2001US-0278152P.
26-MAR-2001; 2001US-0278894P.
27-MAR-2001; 2001US-0279036P.
27-MAR-2001; 2001US-0279036P.
28-MAR-2001; 2001US-0279344P.
30-MAR-2001; 2001US-0279388P.
30-MAR-2001; 2001US-0279959P.
30-MAR-2001; 2001US-0280233P.
02-APR-2001; 2001US-0280802P.
02-APR-2001; 2001US-0280900P.
02-APR-2001; 2001US-0281194P.
04-APR-2001; 2001US-0281675P.
13-APR-2001; 2001US-0283675P.
30-APR-2001; 2001US-0287424P.
02-MAY-2001; 2001US-0288066P.
03-MAY-2001; 2001US-0288342P.
03-MAY-2001; 2001US-0288528P.
15-MAY-2001; 2001US-0291180P.
16-MAY-2001; 2001US-0291099P.
16-MAY-2001; 2001US-0291240P.
30-MAY-2001; 2001US-0294485P.
31-MAY-2001; 2001US-0294889P.
31-MAY-2001; 2001US-0294899P.
18-JUN-2001; 2001US-0299027P.
19-JUN-2001; 2001US-0299303P.
19-JUN-2001; 2001US-0299310P.
10-JUL-2001; 2001US-0303549P.
31-JUL-2001; 2001US-0309198P.
16-AUG-2001; 2001US-0312903P.
10-SEP-2001; 2001US-0318462P.
12-SEP-2001; 2001US-0318770P.
27-SEP-2001; 2001US-0325430P.
27-SEP-2001; 2001US-0325681P.
18-OCT-2001; 2001US-0330380P.
31-OCT-2001; 2001US-0335301P.
14-NOV-2001; 2001US-0332172P.
14-NOV-2001; 2001US-0332271P.
14-NOV-2001; 2001US-0332272P.
14-NOV-2001; 2001US-0333184P.
14-NOV-2001; 2001US-0333272P.

21-NOV-2001; 2001US-0332094P.
03-DEC-2001; 2001US-0337426P.
03-DEC-2001; 2001US-0338092P.
04-DEC-2001; 2001US-0337185P.
03-JAN-2002; 2002US-0345705P.
07-MAR-2002; 2002US-00092900.
(CURA-) CURAGEN CORP.
Padigaru M, Spytek KA, Shenoy SG, Taupier RJ, Pena CE, Li L,
Zerhusen BD, Gusev V, Ji W, Gorman L, Miller CE, Kekuda R,
Paturnajan M, Gangoli E, Vernet CAM, Guo X, Tchernov V,
Fernandes ER, Casman SJ, Malyankar UM, Gerlach V, Liu Y, Anderson D,
Spaderma SK, Catterton E, Burgess C, Lette M, Zhong H, Alsobrook JP,
Lepley DM, Rieger DK,
WPI; 2002-723332/78.
N-PSDB; ABX97157.
NOVX polypeptides and polynucleotides, useful for preventing or treating
a disorder associated with aberrant NOVX expression or activity e.g.,
cancer, hypertension, atherosclerosis, cardiomyopathy or bronchial
asthma.
Claim 1; Page 410; 1103pp; English.
This invention describes novel human NOVX polypeptides which have
cytostatic, cardiant, antiarteriosclerotic, antiasthmatic and hypotensive
activity. Pharmaceutical compositions comprising the NOVX proteins or
nucleic acid molecules or NOVX antibodies are useful for preventing or
treating a disorder associated with aberrant NOVX expression or activity
e.g. cancer, hypertension, atherosclerosis, cardiomyopathy or bronchial
asthma. The products of the invention can be used for gene therapy or in
a vaccine. AB065041-AB065218 represent the NOVX polypeptides encoded by
ABX97008-ABX97185
SQ Sequence 403 AA;
Query Match 35.7%; Score 55; DB 5; Length 403;
Best Local Similarity 46.2%; Pred. No. 22;
Matches 12; Conservative 1; Mismatches 13; Indels 0; Gaps 0;
QY 2 RRGAGARGRASCWVALTLMLAVVP 27
Db 118 RRGSLRGAAGLAWRLFLLAATLP 143
RESULT 11
ABBI6658
ID ABBI6658 standard; protein; 78 AA.
AC ABBI6658;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polypeptide SEQ ID NO 5315.
XX
KW Human; nootropic; neuroprotective; cytosolic; dermatological; virucide;
immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
antiParkinsonian; antisticking; antianaemic; antiarthritic; cancer;
antiheumatic; hepatocytic; cerebroprotective; antiinflammatory;
antiallergic; antidiabetic; anticancer; anticonvulsant; antifungal;
antiparasitic; cardiant; immune disorder; cardiovascular disorder;
neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN MO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US001334.
XX

polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 211 AA;

Query Match 34.4%; Score 53; DB 4; Length 211;
Best Local Similarity 34.6%; Pred. No. 22;
Matches 9; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

1 MRGAGARGASMCWALMLAVV 26
157 MRSVEGFSRSLFMCYLLSLVVSAL 182

ABG07764

ABG36249 standard; protein; 346 AA.

ABG36249;

10-DEC-2002 (first entry)

Maize peroxidase protein Zm-POX07.

Peroxidase-like activity; expression cassette; plant defence response; plant disease resistance; plant stalk strength; plant breeding; maize; plant; peroxidase.

Zea mays.

WO200270723-A2.

12-SEP-2002.

17-JAN-2002; 2002WO-US001454.

18-JAN-2001; 2001US-0262595P.
16-JAN-2002; 2002US-00047825.

(PION-) PIONEER HI-BRED INT INC.

Duvick JP, Maddox JR, Navarro Acevedo PA, Simmons CR;
WPI; 2002-723267/78.
N-PSDB; ABS74468.

New maize peroxidase nucleic acid molecule and polypeptide used in expression cassettes for modulating plant defense response, increasing plant disease resistance and increasing plant stalk strength.

Claim 1; Page 87-88; 119pp; English.

The present invention relates to a new nucleic acid molecule that encodes a polypeptide with peroxidase-like activity. The nucleic acid of the invention and all methods are useful for expression cassettes for modulating plant defence response, increasing plant disease resistance and increasing plant stalk strength. They can also be used in methods of selecting or breeding for plants with increased disease resistance. The present amino acid sequence represents a maize peroxidase protein of the invention

Sequence 346 AA;

Query Match 34.4%; Score 53; DB 5; Length 346;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 12; Conservative 2; Mismatches 2; Indels 8; Gaps 2;

3 RGAGARGASMCWALMLAVV 26
15 RGDGA---AAWCM---WVAVV 30

ABG07764

ABG07764 standard; protein; 600 AA.

ABG07764;

13-FEB-2002 (first entry)

Novel human diagnostic protein #7755.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US008631.

31-MAR-2000; 2000US-00540217.
23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Dzmanac RT, Liu C, Tang YT;
WPI; 2001-639362/73.
N-PSDB; AAS71951.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 20; SEQ ID NO 38123; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 600 AA;

Query Match 34.4%; Score 53; DB 4; Length 600;
Best Local Similarity 66.7%; Pred. No. 63;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

10 GRASMCWALML 21

db 321 GAASWCMVSA 332

Search completed: June 8, 2004, 14:03:56
Job time: 8.42778 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: June 8, 2004, 14:00:33; Search time 2.17778 Seconds
(without alignments)
663.762 Million cell updates/sec

file: US-10-010-050A-2_COPY_1_28

Effect score: 154

quence: 1 MRRCAGAGRGASGCMALMLAVVPG 28

oring table: BLOSUM62

atched: Gapop 10.0, Gapext 0.5

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	34.4	74	4	US-09-621-976-6090
2	52.5	34.1	575	1	US-07-683-957B-2
3	52	33.8	170	4	US-09-252-991A-16705
4	52	33.8	217	4	US-09-252-991A-19862
5	51	33.1	85	4	US-09-431-705-22
6	51	33.1	88	4	US-09-205-258-274
7	51	33.1	1310	4	US-09-170-496D-290
8	51	33.1	1310	4	US-09-364-425B-55
9	50.5	32.6	209	4	US-08-311-731A-18
10	50	32.5	127	4	US-09-252-991A-26303
11	50	32.5	286	4	US-09-252-991A-27451
12	49.5	32.1	92	4	US-09-247-155-98
13	49.5	32.1	196	4	US-09-252-991A-31079
14	49	31.8	718	4	US-09-252-991A-25696
15	49	31.8	1498	4	US-09-252-991A-31234
16	48.5	31.5	251	4	US-09-252-991A-27585
17	48.5	31.5	534	4	US-09-252-991A-17265
18	48	31.2	35	4	US-09-148-545-204
19	48	31.2	170	4	US-09-252-991A-26306
20	48	31.2	215	3	US-09-148-545-135
21	48	31.2	222	4	US-09-252-991A-22195
22	48	31.2	229	4	US-09-252-991A-22195
23	48	31.2	246	4	US-09-252-991A-25102
24	48	31.2	337	4	US-09-252-991A-19602
25	47	30.5	112	4	US-09-252-991A-23296
26	47	30.5	101	4	US-09-107-532A-4380
27	47	30.5	172	4	US-09-252-991A-28284

28	47	30.5	177	4	US-09-252-991A-24167	Sequence 24167, A
29	47	30.5	250	4	US-09-252-991A-24075	Sequence 24075, A
30	47	30.5	439	4	US-09-489-039A-13336	Sequence 13336, A
31	46.5	30.2	308	4	US-09-252-991A-18097	Sequence 18097, A
32	46.5	30.2	418	4	US-09-252-991A-17598	Sequence 17598, A
33	46	29.9	191	2	US-08-469-412A-4	Sequence 4, Appl1
34	46	29.9	191	3	US-09-021-715-4	Sequence 28415, A
35	46	29.9	193	4	US-09-252-991A-28415	Sequence 28054, A
36	46	29.9	200	4	US-09-252-991A-28054	Sequence 176, App
37	46	29.9	248	4	US-09-673-395A-176	Sequence 25031, A
38	46	29.9	335	4	US-09-570-856B-15	Sequence 26787, A
39	46	29.9	478	4	US-09-252-991A-25031	Sequence 23246, A
40	46	29.9	560	4	US-09-252-991A-26787	Sequence 16906, A
41	46	29.9	576	4	US-09-252-991A-23246	Sequence 311, Appl
42	45.5	29.5	263	4	US-09-679-686B-10	Sequence 232, App
43	45.5	29.5	270	4	US-09-489-847-311	
44	45.5	29.5	272	4	US-09-489-847-311	
45	45.5	29.5	273	4	US-09-489-847-232	

ALIGNMENTS

RESULT 1

US-09-621-976-6090

Sequence 6090, Application US/09621976

Patent No. 6639663

GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, J.B.

APPLICANT: Jobert, S.

APPLICANT: Giordano, J.Y.

TITLE OF INVENTION: ESTs and Encoded Human Proteins.

FILE REFERENCE: GENSET.054PR2

CURRENT APPLICATION NUMBER: US/09/621.976

NUMBER OF SEQ ID NOS: 19335

SOFTWARE: Patent .pm

SEQ ID NO 6090

LENGTH: 74

TYPE: PRT

ORGANISM: Homo sapiens

US-09-621-976-6090

Query Match 34.4%, Score 53; DB 4; Length 74;

Best Local Similarity 39.3%; Pred. No. 1.6; Matches 11; Conservative 3; Mismatches 10; Indels 4; Gaps 1;

QY 4 GAGAGGAA-----SCWALALMLAVVP 27

DB 13 GLGAPRGEGPGGQGMCMARSSIFRCYPP 40

RESULT 2

US-07-683-957B-2

Sequence 2, Application US/07683957B

Patent No. 5310880

GENERAL INFORMATION:

APPLICANT: Donahoe, Patricia K.

APPLICANT: Maglin, Richard C.

APPLICANT: MacLaughlin, David T.

TITLE OF INVENTION: Purification of M llerian Inhibiting

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/683,957B
FILING DATE: 19910412
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Goldstein, Jorge A.
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 0609.3060000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 466-0800
TELEFAX: (202) 833-8716
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-683-957B-2

Query Match 34.1%; Score 52.5; DB 1; Length 575;
Best Local Similarity 44.0%; Pred. No. 15;
Matches 11; Conservative 4; Mismatches 7; Indels 3; Gaps 1;

4 GAGAGRGRASWCALMLMLAVPG 28
41 GSGALIFQAWDPISLWL--PG 62

RESULT 3
US-09-252-991A-16705
Sequence 16705; Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 16705
LENGTH: 170
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16705

Query Match 33.8%; Score 52; DB 4; Length 170;
Best Local Similarity 39.3%; Pred. No. 5.1;
Matches 11; Conservative 1; Mismatches 8; Indels 8; Gaps 1;

3 RGAGAGRGRASW-----CMALMLI 22
108 RAAGATRRRPGWTCGANPRDWSAAGSW 135

RESULT 4
US-09-252-991A-19862
Sequence 19862; Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 19862
LENGTH: 217
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19862

Query Match 33.8%; Score 52; DB 4; Length 217;
Best Local Similarity 61.1%; Pred. No. 6.6;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

2 RGAGAGRGRASWCALA 19
191 RRGAGRGGRSGACDGLA 208

RESULT 5
US-09-431-705-22
Sequence 22; Application US/09431705
Patent No. 6585975
GENERAL INFORMATION:
APPLICANT: Kleantous, Harold
APPLICANT: Londono-Arcila, Patricia
APPLICANT: Freeman, Donna
TITLE OF INVENTION: Use of salmonella vectors for
FILE REFERENCE: 06132/060001
CURRENT APPLICATION NUMBER: US/09/431,705
CURRENT FILING DATE: 1999-11-01
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 22
LENGTH: 85
TYPE: PRT
ORGANISM: Escherichia coli
US-09-431-705-22

Query Match 33.1%; Score 51; DB 4; Length 85;
Best Local Similarity 52.9%; Pred. No. 3.5;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

11 RASWCALMLMLAVP 27
16 QASWCARLWMLHSP 32

RESULT 6
US-09-205-258-274
Sequence 274; Application US/09205258
Patent No. 6525174
GENERAL INFORMATION:
APPLICANT: Young et al.
TITLE OF INVENTION: 207 Human Secreted Proteins
FILE REFERENCE: P2007P1
CURRENT APPLICATION NUMBER: US/09/205,258
CURRENT FILING DATE: 1998-12-04
EARLIER APPLICATION NUMBER: PCT/US98/11422
EARLIER FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048,885
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,375
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,881
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,880
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,896
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,020
EARLIER FILING DATE: 1997-06-06

```

/ NAME/KEY: SITE
/ LOCATION: (53)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (88)
/ OTHER INFORMATION: Xaa equals stop translation
US-09-205-258-274

Query Match
Best Local Similarity 33.1%; Score 51; DB 4; Length 88;
Matches 13; Conservative 1; Mismatches 7; Indels 2; Gaps 1;

QY 1 MRGA--GAAGPASCWALAL 21
Db 44 MAKARPEAKRGARWGLAYTL 66

RESULT 7
US-09-170-496D-290
/ Sequence 290, Application US/09170496D
/ Patent No. 6555339
/ GENERAL INFORMATION:
/ APPLICANT: Behan, Dominic P.
/ APPLICANT: Chalmers, Derek T.
/ APPLICANT: Liaw, Chen W.
/ TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-Coupled Receptor
/ FILE REFERENCE: AREN-0040
/ CURRENT APPLICATION NUMBER: US/09/170,496D
/ CURRENT FILING DATE: 1998-10-13
/ NUMBER OF SEQ ID NOS: 294
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 290
/ LENGTH: 1310
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-170-496D-290

Query Match
Best Local Similarity 33.1%; Score 51; DB 4; Length 1310;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 11 RASCWALALMLAVVP 27
Db 907 QASWCHARRLWLHSAP 923

RESULT 8
US-09-364-425B-55
/ Sequence 55, Application US/09364425B
/ Patent No. 6653086
/ GENERAL INFORMATION:
/ APPLICANT: Behan, Dominic P.
/ APPLICANT: Chalmers, Derek T.
/ APPLICANT: Liaw, Chen W.
/ APPLICANT: Lin, I-Lin
/ APPLICANT: Lowitz, Kevin P.
/ APPLICANT: Chen, Rongping
/ TITLE OF INVENTION: Endogenous, Constitutively Activated G Protein-Coupled Orphan Receptor
/ FILE REFERENCE: AREN0047
/ CURRENT APPLICATION NUMBER: US/09/364,425B
/ CURRENT FILING DATE: 2001-12-18
/ PRIOR APPLICATION NUMBER: 60/094,879
/ PRIOR FILING DATE: 1998-07-31
/ PRIOR APPLICATION NUMBER: 60/106,300
/ PRIOR FILING DATE: 1998-10-30
/ PRIOR APPLICATION NUMBER: 60/110,906
/ PRIOR FILING DATE: 1998-12-04
/ PRIOR APPLICATION NUMBER: 60/121,851
/ PRIOR FILING DATE: 1999-02-26
/ NUMBER OF SEQ ID NOS: 60
/ SOFTWARE: PatentIn version 3.1

```

SEQ ID NO 55
LENGTH: 1310
TYPE: PRT
ORGANISM: Homo sapiens
S-09-364-425B-55

Query Match 33.1%; Score 51; DB 4; Length 1310;
Best Local Similarity 52.9%; Pred. No. 59;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Y 11 RASWCWALMLAVVP 27
:|||||:|||||
b 907 QASWCWALMLAVVP 923

RESULT 9
S-08-311-731A-18
Sequence 18, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
TITLES OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 209 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIUM TUBERCULOSIS

JS-08-311-731A-18

Query Match 32.8%; Score 50.5; DB 4; Length 209;
Best Local Similarity 38.7%; Pred. No. 10;
Matches 12; Conservative 2; Mismatches 12; Indels 5; Gaps 1;

Y 1 MRGAGA-----ARGRASWCWALMLAVVP 26
:|||||:|||||
b 163 MRENRSQERVGALTYQWRMSIATITWTTV 193

RESULT 10
JS-09-252-991A-26303
Sequence 26303, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 26303
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26303

Query Match 32.5%; Score 50; DB 4; Length 127;
Best Local Similarity 47.1%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

QY 12 ASWCWALMLAVVP 28
:|||||:|||||
Db 33 ATWCWTLAPTW---Pg 45

RESULT 11
US-09-252-991A-27451
Sequence 27451, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 27451
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27451

Query Match 32.5%; Score 50; DB 4; Length 286;
Best Local Similarity 57.9%; Pred. No. 17;
Matches 11; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

QY 3 RGAGARG---RASWCWA 17
:|||||:|||||
Db 35 RGAGARGDLPFRHGMPWA 53

RESULT 12
US-09-247-155-98
Sequence 98, Application US/09247155A
Patent No. 6312922
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, Jean-Baptiste
APPLICANT: Ducleret, Lydie
TITLE OF INVENTION: Complementary DNAs
FILE REFERENCE: GENSIT.021A
CURRENT APPLICATION NUMBER: US/09/247,155A
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/074,121
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/081,563

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - protein search, using sw model

on on: June 8, 2004, 14:04:04 ; Search time 5.75556 Seconds

(without alignments)
1368.675 Million cell updates/sec

title: US-10-010-050A-2_COPY_1_28
effect score: 154

sequence: 1 MRGGAARGRASWCALALMLAVPG 28

scoring table: BLOSUM62

gapped: Gapped 10.0 , Gapext 0.5

searched: 1155919 seqs, 281338677 residues

total number of hits satisfying chosen parameters: 1155919

minimum DB seq length: 0
maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	346	9	US-09-122-383-2
2	154	100.0	346	12	US-10-653-595-126
3	154	100.0	346	13	US-10-010-050A-2
4	154	100.0	347	12	US-09-397-945-126
5	60.5	39.3	192	9	US-09-811-284-167
6	59.5	38.6	127	15	US-10-291-265-269
7	59	38.3	588	14	US-10-156-761-11538
8	58	37.7	188	9	US-09-989-920-240
9	57	37.0	19608	15	US-10-084-846A-8
10	56	36.4	214	14	US-10-080-170-198
11	55	35.7	403	12	US-10-092-900A-300
12	53	34.4	160	14	US-10-156-761-10933
13	53	34.4	346	12	US-10-047-825-10
14	53	34.4	368	12	US-10-425-114-55731
15	52.5	34.1	199	12	US-10-221-278-727

16	52.5	34.1	199	15	US-10-291-172-727	Sequence 727, App
17	52.5	34.1	333	9	US-09-738-626-4771	Sequence 4771, App
18	52.5	34.1	360	12	US-10-282-122A-50396	Sequence 50396, A
19	52.5	34.1	464	12	US-10-424-599-207280	Sequence 207280,
20	52	33.8	82	12	US-10-424-599-167251	Sequence 167251,
21	52	33.8	560	10	US-09-863-776-69	Sequence 69, Appl
22	52	33.8	599	15	US-10-104-047-2306	Sequence 2306, Ap
23	52	33.8	653	10	US-09-863-776-28	Sequence 28, Appl
24	52	33.8	658	9	US-09-759-056-5	Sequence 5, Appl
25	52	33.8	658	9	US-09-901-812-5	Sequence 5, Appl
26	52	33.8	658	10	US-09-863-776-68	Sequence 68, Appl
27	52	33.8	662	10	US-09-863-776-30	Sequence 30, Appl
28	52	33.8	667	9	US-09-759-056-2	Sequence 2, Appl
29	52	33.8	667	9	US-09-901-812-2	Sequence 2, Appl
30	52	33.8	667	10	US-09-863-776-32	Sequence 32, Appl
31	52	33.8	667	10	US-09-863-776-67	Sequence 67, Appl
32	52	33.8	667	12	US-10-219-535-80	Sequence 80, Appl
33	52	33.8	667	12	US-10-232-230-80	Sequence 80, Appl
34	52	33.8	667	12	US-10-232-224-80	Sequence 80, Appl
35	52	33.8	667	14	US-10-237-884-80	Sequence 80, Appl
36	52	33.8	667	14	US-10-230-163-80	Sequence 80, Appl
37	52	33.8	667	14	US-10-230-338-80	Sequence 80, Appl
38	52	33.8	667	14	US-10-218-631-80	Sequence 80, Appl
39	52	33.8	667	14	US-10-230-414-80	Sequence 80, Appl
40	52	33.8	667	14	US-10-216-159A-80	Sequence 80, Appl
41	52	33.8	667	14	US-10-218-849-80	Sequence 80, Appl
42	52	33.8	667	14	US-10-227-873-80	Sequence 80, Appl
43	52	33.8	667	14	US-10-227-883-80	Sequence 80, Appl
44	52	33.8	667	14	US-10-219-076-80	Sequence 80, Appl
45	52	33.8	667	14	US-10-230-434-80	Sequence 80, Appl

ALIGNMENTS

RESULT 1
US-09-122-383-2
Sequence 2, Application US/09122383A
Patent No. US20020042093A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Gilbertson, Debra G.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
FILE REFERENCE: 97-38
CURRENT APPLICATION NUMBER: US/09/122,383A
CURRENT FILING DATE: 1998-07-24
EARLIER APPLICATION NUMBER: 60/053,613
EARLIER FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 346
TYPE: PRT
ORGANISM: Homo sapien
US-09-122-383-2

Query Match 100.0%; Score 154; DB 9; Length 346;
Best Local Similarity 100.0%; Pred. No. 6e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRGGAARGRASWCALALMLAVPG 28
|||
Db 1 MRGGAARGRASWCALALMLAVPG 28

RESULT 2
US-10-653-595-126
Sequence 126, Application US/10653595
Publication No. US20040048304A1
GENERAL INFORMATION:
APPLICANT: Ruben et. al.
TITLE OF INVENTION: 95 Human secreted proteins

FILE REFERENCE: PZ027P1C1
CURRENT APPLICATION NUMBER: US/10/653,595
CURRENT FILING DATE: 2003-09-03 09:397945
PRIOR APPLICATION NUMBER: US 09/397945
PRIOR FILING DATE: 1999-09-17
PRIOR APPLICATION NUMBER: PCT/US99/05804
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 126
LENGTH: 346
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (242)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (246)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-653-595-126
Query Match 100.0%; Score 154; DB 12; Length 346;
Best Local Similarity 100.0%; Pred. No. 6e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2y 1 MRGGAAGRGRASWCWALALMLAVPG 28
Db 1 MRGGAAGRGRASWCWALALMLAVPG 28
RESULT 3
US-10-010-050A-2
Sequence 2, Application US/10010050A
Publication No. US20020173624A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Gilbertson, Debra G.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
FILE REFERENCE: 97-38C1
CURRENT APPLICATION NUMBER: US/10/010,050A
CURRENT FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: US 09/122,383
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: US 60/053,613
PRIOR FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 346
TYPE: PRT
ORGANISM: Homo sapien
US-10-010-050A-2

Query Match 100.0%; Score 154; DB 13; Length 346;
Best Local Similarity 100.0%; Pred. No. 6e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MRGGAAGRGRASWCWALALMLAVPG 28
Db 1 MRGGAAGRGRASWCWALALMLAVPG 28
RESULT 4
US-09-397-945-126
Sequence 126, Application US/09397945
Publication No. US20030065139A1
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc. et al.
TITLE OF INVENTION: 95 Human secreted proteins
FILE REFERENCE: PZ027P1
CURRENT APPLICATION NUMBER: US/09/397,945
CURRENT FILING DATE: 1999-09-17
PRIOR APPLICATION NUMBER: PCT/US99/05804
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,581
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,577
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,563
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,313
PRIOR FILING DATE: 1998-04-01
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 126
LENGTH: 347
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (242)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (246)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (347)
OTHER INFORMATION: Xaa equals stop translation
US-09-397-945-126
Query Match 100.0%; Score 154; DB 12; Length 347;
Best Local Similarity 100.0%; Pred. No. 6e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MRGGAAGRGRASWCWALALMLAVPG 28
Db 1 MRGGAAGRGRASWCWALALMLAVPG 28
RESULT 5

```

S-09-811-284-167
Sequence 167, Application US/09811284
Patent No. US20020058306A1

GENERAL INFORMATION:
APPLICANT: Vogeli, Gabriel
TITLE OF INVENTION: No. US20020058306A1e1 G Protein-Coupled Receptors
FILE REFERENCE: 00167US1
CURRENT APPLICATION NUMBER: US/09/811,284
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/189,783
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,907
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,918
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,960
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,917
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/192,945
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,916
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,923
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,933
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,830
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,234
PRIOR FILING DATE: 2000-03-27
PRIOR APPLICATION NUMBER: 60/192,155
PRIOR FILING DATE: 2000-03-24
PRIOR APPLICATION NUMBER: 60/192,935
PRIOR FILING DATE: 2000-03-29
NUMBER OF SEQ ID NOS: 258
SOFTWARE: PatentIn version 3.0

SEQ ID NO 167
LENGTH: 192
TYPE: PRT
ORGANISM: Homo sapiens
S-09-811-284-167

Query Match          39.3%;   Score 60.5;   DB 9;   Length 192;
Best local similarity 39.3%;   Pred. No. 8.3;
Matches 11;   Conservative 2;   Mismatches 6;   Indels 9;   Gaps 1.

      4  GAGARGASWCW-----ALALIM 22
      ||| ||| |||
1b  35  GAAVGRPCWPCPCPALVSTALW 62

RESULT 6
Sequence 269, Application US/10291265
Publication No. US20030232054A1

GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
APPLICANT: Tang et al
TITLE OF INVENTION: No. US20030232054A1e1 Nucleic Acids and Polypeptides
FILE REFERENCE: 21272-017 (785)
CURRENT APPLICATION NUMBER: US/10/291,265
CURRENT FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: 09/617,746
PRIOR FILING DATE: 2000-07-17
PRIOR APPLICATION NUMBER: 09/631,451
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 09/633,870
PRIOR FILING DATE: 2000-09-15
NUMBER OF SEQ ID NOS: 944
SOFTWARE: FastSeq for Windows Version 3.0

```

```

; SEQ ID NO 269
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-269

Query Match
Best Local Similarity 38.6%; Score 59.5; DB 15; Length 127;
Matches 14; Conservative 3; Mismatches 9; Indels 9; Gaps 2;

QY      3   RGAGAAAGR-----ASNCWALAIL--WLAVVPG 28
          |||:|||||::|||::|||:
Db       25   RGEGETGCVTMTKLAQMLWGIALIGSTVALTTG 59

RESULT 7
US-10-156-761-11538
; Sequence 11538, Application US/10156761
; Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 11538
LENGTH: 588
TYPE: PRT
ORGANISM: Streptomyces avermiltilis
US-10-156-761-11538

Query Match
Best Local Similarity 38.3%; Score 59; DB 14; Length 588;
Matches 13; Conservative 5; Mismatches 4; Indels 22; Gaps 1;

QY      3   RGAGAAR-----GRASWCWALLLWMA 24
          |||:|||||::|||::|||:
Db       167   RGAGSSRTKYATKNLTLLGYGSRHYGGDKSWSNAGGLTWLA 210

RESULT 8
US-09-989-920-240
; Sequence 240, Application US/09989920
; Patent No. US20020172957A1
GENERAL INFORMATION:
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Hervé
APPLICANT: Chen, Sei-Yu
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenchua
TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
FILE REFERENCE: DEX-0291
CURRENT APPLICATION NUMBER: US/09/989,920
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/252,500
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 284
SOFTWARE: PatentIn version 3.1
SEQ ID NO 240
LENGTH: 188
TYPE: PRT
ORGANISM: Homo sapien

```


Query Match 34.1%; Score 52.5; DB 12; Length 199;
Best Local Similarity 73.3%; Pred. No. 81;
Matches 11; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

ly 2 RRGAGAGRGRASWCW 16
| | | | | | | | | |
yb 164 RDGAGRAGRRA-WYW 177

Search completed: June 8, 2004, 14:11:57
Job time : 6.75556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

4 protein - protein search, using sw model

on: June 8, 2004, 13:59:38 ; Search time 2.02222 Seconds
(without alignments)
1331.683 Million cell updates/sec

File: US-10-010-050a-2_COPY_1_28
Sequence: 1 MRGAGARGASMCWALALIMLAVVPG 28

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 283366 seqs, 96191526 residues

otal number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

set-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	42.9	606	2 C84390	sulfate transport
2	59	38.3	412	2 D95941	conserved hypothet
3	56	36.4	214	2 T44701	probable integral
4	54	35.1	232	2 F82729	conserved hypothet
5	53	34.4	428	2 A12860	conserved hypothet
6	53	34.4	428	2 H97637	hypothetical prote
7	52.5	34.1	575	1 WPROM	multimeric inhibiti
8	52	33.8	641	2 A45054	probable intercell
9	51.5	33.4	686	2 S30075	ferric reductase (
10	51.5	33.4	1146	2 A38587	collagen, cornea-s
11	51	33.1	410	2 E95984	conserved hypothet
12	50.5	32.8	210	2 E70671	hypothetical prote
13	50	32.5	106	2 A29760	hypothetical prote
14	50	32.5	163	2 S73039	hypothetical prote
15	50	32.5	221	2 G83964	hypothetical prote
16	50	32.5	249	2 B84147	ABC transporter (p
17	50	32.5	250	2 AD1197	ABC transporter tr
18	50	32.5	417	2 AF3448	hypothetical membr
19	49.5	32.1	201	2 D81944	probable integral
20	49	31.8	118	2 S52855	hypothetical prote
21	49	31.8	139	2 A71123	hypothetical prote
22	49	31.8	188	2 B82183	ank protein VC158
23	49	31.8	250	2 AB1555	ABC transporter tr
24	49	31.8	310	2 T16233	hypothetical prote
25	49	31.8	317	2 T35010	probable integral
26	49	31.8	755	2 B75346	probable compen
27	49	31.8	1221	2 B83327	conserved hypothet
28	48.5	31.5	210	2 S36297	T-cell receptor ga
29	48.5	31.5	435	2 A12082	hypothetical prote

30	48.5	31.5	519	2 S75570	apolipoprotein N-a
31	48.5	31.5	688	2 B97152	probable membrane
32	48.5	31.5	840	2 AG0526	penicillin-binding
33	48.5	31.5	854	2 D82324	conserved hypothet
34	48	31.2	159	2 A48428	homeotic protein H
35	48	31.2	221	2 UC7587	stromal cell-deriv
36	48	31.2	233	2 B40364	hypothetical prote
37	48	31.2	334	2 F75539	branched-chain am
38	48	31.2	365	2 A82941	hypothetical prote
39	48	31.2	369	2 H98341	probable di-trans,
40	48	31.2	400	2 T35334	probable membrane
41	48	31.2	968	2 T00322	hypothetical prote
42	48	31.2	1245	1 MMSND	nidogen precursor
43	47.5	30.8	519	2 F96517	hypothetical prote
44	47.5	30.8	737	2 H66538	hypothetical prote
45	47.5	30.8	1360	2 JCS839	GTBR-N protein - h

ALIGNMENTS

RESULT 1

C84390 sulfate transport system permease protein [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84390
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laaky, S.; Lechauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lie A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: C84390
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-606 <STO>
A:Cross-references: GB:AB004437; NID:g10581803; PID:AA020487.1; GSPDB:GN00138
A:Gene: cyst1

Query Match 42.9%; Score 66; DB 2; Length 606;
Best Local Similarity 53.8%; Pred. No. 0.31;
Matches 14; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 3 RGAGARGASMCWALALIMLAVVPG 28
DB 508 RALGARGALMDVPLVPGVAVG 533

RESULT 2

D95941 conserved hypothetical membrane protein, paralogue of Y20848 SMB21292 [imported] - Sinorhizobium meliloti
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: D95941
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmeier, J.; Chain, P.; Vorholter, F.J.; Hernan Proc. Natl. Acad. Sci. U.S.A. 98, 9869-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N₂-fixing endo A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: D95941
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <KUR>
A:Cross-references: GB:AL591985; PID:CA049196.1; PID:g15140681; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, A.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C.

;Molecule type: DNA
;Residues: 1-221 <STO>
;Cross-references: GB:AP001515; GB:BA000004; NID:G10174886; PIDN:BA06238.1; GSPDB:GNOC
;Experimental source: strain C-125
;Genetics:
;Gene: BH2519

Query Match 32.5%; Score 50; DB 2; Length 221;
Best Local Similarity 41.7%; Pred. No. 18;
Matches 10; Conservative 2; Mismatches 12; Indels 0; Gaps 3;

Y 4 GAGAGRGRASWCVALALLMLAVVP 27
D 31 GIGALIGAAFPWHLAWAILFP 54

Search completed: Tue 8, 2004, 14:08:08
3D time : 3.02222 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

4 protein - protein search, using sw model

on: June 8, 2004, 13:56:38 ; Search time 1.16667 Seconds

(without alignments)
1249.684 Million cell updates/sec

file: US-10-010-050a-2_COPY_1_28
sequence: 1 MRRGAGARGRASCMALALMLAVVPG 28

scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

sarched: 141681 seqs, 52070155 residues

cal number of hits satisfying chosen parameters: 141681

imum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	407	1	CLN5 HUMAN
2	56.5	36.7	126	1	CRGB RALSO
3	52.5	34.1	575	1	MIS BOVIN
4	52	33.8	641	1	FZD1 RAT
5	51.5	33.4	686	1	FRE1 YEAST
6	50	32.5	106	1	SH RAT
7	49.5	32.1	92	1	DEP3 HUMAN
8	49	31.8	286	1	CLN8 HUMAN
9	49	31.8	330	1	PEXD CAEEL
10	48.5	31.5	519	1	LNT SYNY3
11	48	31.2	162	1	HXA9 CAVPO
12	48	31.2	221	1	SDPL HUMAN
13	48	31.2	233	1	PIC1 AGRT5
14	48	31.2	1245	1	NIDO MOUSE
15	47	30.5	339	1	YJG8 ECOLI
16	47	30.5	343	1	GAS1 MOUSE
17	47	30.5	345	1	GAS1 HUMAN
18	47	30.5	398	1	GSP1 KLEPN
19	46	29.9	508	1	Y202 HUMAN
20	46	29.9	335	1	XVNB STRLI
21	46	29.9	812	1	UBA4 HUMAN
22	45.5	29.5	108	1	YNPA SALTY
23	45.5	29.5	249	1	MSX1 CHICK
24	45.5	29.5	844	1	PBPB ECOLI
25	45.5	29.5	1112	1	CN3B HUMAN
26	45.5	29.5	2832	1	NDBV RHIME
27	45	29.2	359	1	CD24 MOUSE
28	45	29.2	362	1	CNTR CHICK
29	45	29.2	755	1	LEB2 RHIPU
30	45	29.2	1021	1	CARA MOUSE
31	45	29.2	1041	1	CHS1 CRYNE
32	45	29.2	1968	1	RPO PVMR
33	44.5	28.9	92	1	DEP3 MOUSE

34	44.5	28.9	108	1	YNPA SALTY	O82622 salmonella
35	44.5	28.9	307	1	RCGM RHOSH	P02853 rhodobacter
36	44.5	28.9	420	1	MLIC XENLA	P49219 xenopus lae
37	44.5	28.9	451	1	NORM DEIRA	Q94744 dehnococcus
38	44.5	28.9	482	1	CLS2 BACSV	P71040 bacillus su
39	44.5	28.9	758	1	CSTA MYCTU	P95095 mycobacteri
40	44.5	28.9	1356	1	HEM1 PODAN	O00808 podospora a
41	44	28.6	332	1	OSTY YEAST	O03723 saccharomyc
42	44	28.6	442	1	TM14 HUMAN	Q14142 homo sapien
43	44	28.6	459	1	SUXO CHICK	P07850 gallus gall
44	44	28.6	462	1	HEMO HUMAN	P02790 homo sapien
45	44	28.6	560	1	FXJB MOUSE	P59997 mus musculu

ALIGNMENTS

```

RESULT 1
ID CLN5_HUMAN STANDARD; PRT; 407 AA.
AC 075503;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ceroid-lipofuscinosis neuronal protein 5 (CLN5 protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., VARIANT VLINCL ASN-279, AND VARIANT ARG-368.
RC TISSUE=Petal brain;
RX MEDLINE=98324783; PubMed=9662406;
RA Savukoski M., Kloockars T., Holmberg V., Santavuori P., Lander E.S.,
RA Peltonen L.;
RT "CLN5, a novel gene encoding a putative transmembrane protein mutated
RT in Finnish variant late infantile neuronal ceroid lipofuscinosis.";
RN Nat. Genet. 19:286-288 (1998).
[2]
SUBCELLULAR LOCATION, AND GLYCOSYLATION.
MEDLINE=21968572; PubMed=11971870;
RA Isoomppi J., Vesa J., Jalanko A., Peltonen L.;
RT "Lysosomal localization of the neuronal ceroid lipofuscinosis CLN5
RT protein.";
RL Hum. Mol. Genet. 11:885-891 (2002).
CC -!- FUNCTION: Not known.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- PTM: Glycosylated.
CC -!- DISEAS: Defects in CLN5 are the cause of Finnish variant late-
CC infantile neuronal ceroid lipofuscinosis (VLINCL) [MIM:256731];
CC also known as ceroid lipofuscinosis neuronal 5 (CLN5). VLINCL is a
CC fatal childhood neurodegenerative disease characterized by
CC progressive visual and mental decline, motor disturbance, epilepsy
CC and behavioral changes. The first symptom is motor clumsiness,
CC followed by progressive visual failure, mental and motor
CC deterioration and later by myoclonia and seizures.
CC -!- DATABASE: NAME=NCL CLN5;
CC NOTE=Neural Ceroid Lipofuscinoses mutation db;
CC WWW="http://www.ucl.ac.uk/ncl/CLN5.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF068227; AAC27614.1; -
CC DR Genew; HGNC:2076; CLN5.
CC DR MIM; 608102; -.

```


JR MIM; 256731; -
 JR GO: GO:0016021; C: integral to membrane; TAS.
 JR GO: GO:0008151; P: cell growth and/or maintenance; TAS.
 JR Transmembrane; Lysosome; Glycoprotein; Neuronal ceroid lipofuscinosis;
 CM Disease mutation; Polymorphism; Epilepsy.
 TR TRANSMEM 75
 TR CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
 TR CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
 TR CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
 TR CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 TR CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
 TR CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 TR CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 TR CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
 TR VARIANT 279 279 D -> N (in VLIINCL).
 TR VARIANT 368 368 /FTID=VAR_005137.
 TR /FTID=VAR_005138.
 TR SEQUENCE 407 AA; 46339 MM; 449702DIDC9BFEE4 CRC64;
 Query Match 100.0%; Score 154; DB 1; Length 407;
 Best Local Similarity 100.0%; Pred. No. 3.5e-13;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRGAGAGRAGRAGCAGATATLMLAVPG 28
 62 MRGAGAGRAGRAGCAGATATLMLAVPG 89

RESULT 2

JC CBH_RALSO STANDARD; PRT: 126 AA.
 JC Q8XZR2; 28-FEB-2003 (Rel. 41, Created)
 JC 28-FEB-2003 (Rel. 41, Last sequence update)
 JC 28-FEB-2003 (Rel. 41, Last annotation update)
 JC Protein cRcb homolog.
 JC CRCB OR RSC133 OR RS02855.
 JC Ralstonia solanacearum (Pseudomonas solanacearum).
 JC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 JC Burkholderiaceae; Ralstonia.
 JC NCBI TaxID=305;
 JC [1]
 JC SEQUENCE FROM N.A.
 JC MEDLINE=21681879; PubMed=11823852;
 JC Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 JC Ariat M., Billault A., Broctier P., Camus J.C., Catolico L.,
 JC Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
 JC Gaupin C., Lavie M., Moisan A., Robert C., Saurin W., Schiek T.,
 JC Siguiet P., Thebaud P., Whalen M., Wincker P., Levy M.,
 JC Weissenbach J., Boucher C.A.;
 JC "Genome sequence of the plant pathogen Ralstonia solanacearum";
 JC Nature 415:497-502(2002).
 JC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 JC -1- SIMILARITY: Belongs to the cRcb family.
 JC -----
 JC This SWISS-PROT entry is copyright. It is produced through a collaboration
 JC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 JC the European Bioinformatics Institute. There are no restrictions on its
 JC use by non-profit institutions as long as its content is in no way
 JC modified and this statement is not removed. Usage by and for commercial
 JC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 JC or send an email to license@isb-sib.ch).
 JC -----
 JC EMBL; AF646064; CAD15035.1; -
 JR HAMAP; MF 00454; -; 1.
 JR InterPro; IPR003691; Campor_Crcb.
 JR Pfam; PF02537; CRCB; 1.
 JR TIGRFAMs; TIGR00494; cRcb; 1.
 JR Transmembrane; Complete proteome.
 JR TRANSMEM 5 24 POTENTIAL.
 JR TRANSMEM 34 56 POTENTIAL.

FT TRANSMEM 68 90 POTENTIAL.
 FT TRANSMEM 100 122 POTENTIAL.
 SO SEQUENCE 126 AA; 13101 MM; C2443FBA5C81CB3 CRC64;
 Query Match 36.7%; Score 56.5; DB 1; Length 126;
 Best Local Similarity 52.0%; Pred. No. 0.85;
 Matches 13; Conservative 3; Mismatches 6; Indels 3; Gaps 2;

4 GAGAGRAGRAGCAGATATLMLAVPG 27
 10 GAGAGRAGRAGCAGATATLMLAVPG 32

RESULT 3

MTS_BOVIN
 ID MTS_BOVIN STANDARD; PRT: 575 AA.
 AC P03972;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Muellerian inhibiting factor precursor (MTS) (Anti-muellerian hormone)
 DE (AMH) (Muellerian inhibiting substance).
 GN AMH.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI TaxID=9913;
 RX [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP MEDLINE=86218082; PubMed=3754790;
 RA Cate R.L., Mataliano R.J., Hession C., Tizard R., Farber N.M.,
 RA Cheung A., Nifra E.G., Frey A.Z., Gash D.J., Chow E.P., Fisher R.A.,
 RA Bertonis J.W., Torres G., Wallner B.P., Ramachandran K.L.,
 RA Rabin R.C., Mangano T.F., McLaughlin D.T., Donahoe P.K.;
 RT "Isolation of the bovine and human genes for Mullerian inhibiting
 RT substance and expression of the human gene in animal cells";
 RT Cell 45:685-698(1986).
 RL -1- FUNCTION: THIS GLYCOPROTEIN, PRODUCED BY THE SEROTOL CELLS OF THE
 CC TESTIS, CAUSES REGRESSION OF THE MUELLERIAN DUCT. IT ALSO IS ABLE
 CC TO INHIBIT THE GROWTH OF TUMORS DERIVED FROM TISSUES OF MUELLERIAN
 CC DUCT ORIGIN.
 CC -1- SUBUNIT: Homodimer; disulfide-linked.
 CC -1- MISCELLANEOUS: Although it does not compete with EGF for receptor
 CC binding sites, MTS can inhibit the autophosphorylation of the EGF
 CC receptor in vitro.
 CC -1- SIMILARITY: Belongs to the TGF-beta family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M13151; AAA98765.1; -
 JR PIR; A01398; MEBOM.
 JR InterPro; IPR006799; AMH_N.
 JR InterPro; IPR002400; GF_cysknot.
 JR InterPro; IPR001839; TGFb.
 JR Pfam; PF04709; AMH_N; 1.
 JR Pfam; PF00019; TGF-beta; 1.
 JR PRINTS; PR00439; GRCYSKNOT.
 JR PRODOM; PD000357; TGFb; 1.
 JR SMART; SM00204; TGFb; 1.
 JR PROSITE; PS00250; TGF_BETA_1; 1.
 JR Growth factor; Glycoprotein; Gonadal differentiation; Signal.
 FT SIGNAL 1 17
 FT PROPEP 19 24
 FT CHAIN 25 575
 FT DISULFID 477 541
 FT DISULFID 503 572
 FT BY SIMILARITY.

```

T DISUFIID 507 574 BY SIMILARITY.
T DISUFIID 540 540 INTERCHAIN (BY SIMILARITY).
T CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
Q SEQUENCE 575 AA; 60623 MW; 8928B9C11ACB5A8 CRC64;

Query Match 34.1%; Score 52.5; DB 1; Length 575;
Best Local Similarity 44.0%; Pred. No. 11;
Matches 11; Conservative 4; Mismatches 7; Indels 3; Gaps 1;

Y 4 GAGARGRASCWALALMLAVVPG 28
b 41 GSGALIFQAMWPLSLW---PG 62

RESULT 4
ZDI_RAT STANDARD; PRT; 641 AA.
ZDI_RAT
008463;
28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Frizzled 1 precursor (Frizzled-1) (Fz-1) (Fz1).
FZD1.
Rattus norvegicus (Rat).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
(1)
SEQUENCE FROM N.A.
STRAIN=Sprague-Dawley; TISSUE=Osteosarcoma;
MEDLINE=93094228; PubMed=1334084;
Chan S.D.H., Karpf D.B., Fowlkes M.E., Hooks M., Bradley M.S.,
Vuong V., Bambo T., Liu M.Y.C., Arnard C.D., Strewler G.J.,
Nissenen R.A.;
"Two homologs of the Drosophila polarity gene frizzled (fz) are widely
expressed in mammalian tissues.";
J. Biol. Chem. 267:25202-25207(1992).
(2)
COUPLING TO BETA-CATENIN PATHWAY.
MEDLINE=99324245; PubMed=10395542;
Sheldahl L.C., Park M., Malbon C.C., Moon R.T.;
"Protein kinase C is differentially stimulated by Wnt and Frizzled
homologs in a G-protein-dependent manner.";
Curr. Biol. 9:695-698(1999).
-1- FUNCTION: Receptor for Wnt proteins. Most of frizzled receptors
are coupled to the beta-catenin canonical signaling pathway, which
leads to the activation of dishevelled proteins, inhibition of
GSK-3 kinase, nuclear accumulation of beta-catenin and activation
of Wnt target genes. A second signaling pathway involving PKC and
calcium fluxes has been seen for some family members, but it is
not yet clear if it represents a distinct pathway or if it can be
integrated in the canonical pathway, as PKC seems to be required
for Wnt-mediated inactivation of GSK-3 kinase. Both pathways seem
to involve interactions with G-proteins. May be involved in
transduction and intercellular transmission of polarity
information during tissue morphogenesis and/or in differentiated
tissues. Activation by Wnt induces expression of beta-catenin
target genes.
-1- SUBCELLULAR LOCATION: Integral membrane protein.
-1- TISSUE SPECIFICITY: Widely expressed. Most abundant in kidney,
liver, uterus, ovary and heart. Lower levels seen in brain and
intestine. Extremely low in calvaria, mammary glands and testis.
-1- DEVELOPMENTAL STAGE: Expressed predominantly in neonatal tissues,
at lower levels in adult.
-1- DOMAIN: Lys-thr-x-x-x-trp motif is involved in the activation of
the Wnt/beta-catenin signaling pathway (By similarity).
-1- DOMAIN: The Fz domain is involved in binding with Wnt ligands (By
similarity).
-1- SIMILARITY: Belongs to the Fz/Smo G-protein coupled receptor
family.
-1- SIMILARITY: Contains 1 frizzled (FZ) domain.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

DR EMBL; L02529; AAA41173.1; -.
DR InterPro; IPR000539; Frizzled.
DR InterPro; IPR000024; Fz domain.
DR InterPro; IPR000832; GPCR secretin.
DR Pfam; PF01534; Frizzled; 1.
DR Pfam; PF01392; Fz; 1.
DR PRINTS; PRO0489; FRIZZLED.
DR SMART; SM00063; FRI; 1.
DR PROSITE; PS0038; FZ; 1.
DR PROSITE; PS0261; G-PROTEIN RECEPTOR; Transmembrane;
  KM Multigene family; G-protein coupled receptor; Transmembrane;
  KW Developmental protein; Wnt signaling pathway; Glycoprotein; Signal.
FT SIGNAL 1 68
FT CHAIN 1 641
FT DOMAIN 69 316
FT TRANSMEM 317 337
FT DOMAIN 338 348
FT TRANSMEM 349 369
FT DOMAIN 370 396
FT TRANSMEM 397 417
FT DOMAIN 418 439
FT TRANSMEM 440 460
FT DOMAIN 461 483
FT TRANSMEM 484 504
FT TRANSMEM 505 530
FT TRANSMEM 531 551
FT DOMAIN 552 595
FT TRANSMEM 596 616
FT DOMAIN 617 641
FT DOMAIN 106 224
FT SITE 85 90
FT SITE 619 624
FT SITE 639 641
FT CARBOHYD 125 125
FT CARBOHYD 225 225
SQ SEQUENCE 641 AA; 71027 MW; D8E2C113B818B6 CRC64;

```

Query Match 33.8%; Score 52; DB 1; Length 641;
 Best Local Similarity 55.6%; Pred. No. 14;
 Matches 15; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

```

QY 5 AGARGRASCWALALMLAVV 27
Db 38 AGHRRFRASRCWARGLLMLLEAP 64

```

```

RESULT 5
PREL_YEAST STANDARD; PRT; 686 AA.
AC P32791.
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ferric reductase transmembrane component 1 precursor (EC 1.16.1.7)
DE (Ferric-chelate reductase 1).
GN PREL OR YKR214W OR U8167.2.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F113;
RX MEDLINE=92237270; PubMed=1570306;
RA Dancis A., Roman D.G., Anderson G.J., Himebusch A.G., Klausner R.D.;

```

RT "Ferric reductase of Saccharomyces cerevisiae: molecular
 RT characterization, role in iron uptake, and transcriptional control by
 RT iron.";
 RT Proc. Natl. Acad. Sci. U.S.A. 89:3869-3873(1992).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RC MEDLINE=97313267; PubMed=9169871;
 RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansgore W.,
 RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
 RA Enrich K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
 RA Heuss-Weitel D., Hilbert H., Hilger F., Kleine K., Koelter P.,
 RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
 RA Mueller-Auer S., Nentwich U., Obermayer B., Piravandi E., Pohl T.M.,
 RA Portetlelle D., Punelle B., Reckmann S., Rieger M., Rinke M., Rose M.,
 RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz S.,
 RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhaselt P.,
 RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambut R., Wedler E.,
 RA Wedler H., Zimmermann F.K., Zollner A., Hant U., Hohlseil J.D.,
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
 RT Nature 387:87-90(1997).
 RN [3]
 RP CHARACTERIZATION.
 RP MEDLINE=93057491; PubMed=1331884;
 RA Anderson G.J., Jesuisse E., Dancis A., Roman D.G., Tabbe P.,
 RA Klausner R.D.;
 RT "Ferric iron reduction and iron assimilation in Saccharomyces
 RT cerevisiae.";
 RT J. Inorg. Biochem. 47:249-255(1992).
 RC -I- FUNCTION: Reductase activity that acts on ferric iron chelates
 RC external to the cell. Plays a role in iron uptake. May
 RC participate in the transport of electrons from cytoplasm to an
 RC extracellular substrate (ferric ion) via FAD and heme
 RC intermediates. May also participate in Cu(II) reduction and Cu(I)
 RC uptake.
 RC -I- CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).
 RC -I- COFACTOR: FAD (probable).
 RC -I- SUBCELLULAR LOCATION: Integral membrane protein.
 RC -I- INDUCTION: BY IRON DEPRIVATION. REPRESSED BY IRON UPTAKE.
 RC -I- SIMILARITY: Belongs to the FRE / CYB family.
 RC
 RC This SWISS-PROT entry is copyright. It is produced through a collaboration
 RC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 RC the European Bioinformatics Institute. There are no restrictions on its
 RC use by non-profit institutions as long as its content is in no way
 RC modified and this statement is not removed. Usage by and for commercial
 RC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 RC or send an email to license@isb-sib.ch).
 RC
 RC -----
 RC EMBL; M86908; AAA34608.1; -;
 RC EMBL; U14913; AAB67424.1; -;
 RC PIR; S30075; S30075.
 RC Genmonline; 142276; -;
 RC SGD; S0004204; FRE1.
 RC GO; GO:0005886; C:plasma membrane; IDA.
 RC GO; GO:0000293; F:ferric-chelate reductase activity; IDA.
 RC GO; GO:0015677; P:copper ion import; IDA.
 RC GO; GO:0006826; P:iron ion transport; IDA.
 RC InterPro; IPR002916; Ferric_reduct.
 RC Pfam; PF01794; Ferric_reduct; 1.
 RC Oxidoreductase; Electron transport; Transmembrane; Iron transport;
 RC FAD; NAD; Copper; Glycoprotein; Signal; Multigene family.
 RC SIGNAL
 RC CHAIN 1 22
 RC 23 686
 RC 1
 RC FERRIC REDUCTASE TRANSMEMBRANE COMPONENT
 RC
 RC NP_BIND 462 468 FAD (POTENTIAL).
 RC NP_BIND 532 540 NAD (POTENTIAL).
 RC TRANSMEM 147 169 1 (POTENTIAL).
 RC TRANSMEM 216 236 2 (POTENTIAL).
 RC TRANSMEM 258 277 3 (POTENTIAL).
 RC TRANSMEM 296 316 4 (POTENTIAL).
 RC TRANSMEM 329 348 5 (POTENTIAL).
 RC TRANSMEM 369 397 6 (POTENTIAL).

FT TRANSMEM 529 550 7 (POTENTIAL).
 FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 686 AA; 78853 MW; 7P6BB3B93A95D6A3 CRC64;
 Query Match 33.4%; Score 51.5; DB 1; Length 686;
 Best Local Similarity 47.1%; Pred. No. 17;
 Matches 8; Conservative 3; Mismatches 5; Indels 1; Gaps 1;
 Oy 11 RASWC-MALMULTLAVV 26
 Db 144 RSGWCAGLVFVAVL 160
 RESULT 6
 SH_RAT
 ID SH_RAT STANDARD; PRT; 106 AA.
 AC P55248;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Putative protein SH (SH-4).
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TTSUSE-Heart;
 RC MEDLINE=87149087; PubMed=3547652;
 RA Adelman J.P., Bond C.T., Douglas J., Herbert E.;
 RT "Two mammalian genes transcribed from opposite strands of the same
 RT DNA locus.";
 RL Science 235:1514-1517(1987).
 CC -I- FUNCTION: May be involved with the regulation of GNRH gene
 CC expression. It is not known if this protein is transcribed.
 CC -I- TISSUE SPECIFICITY: Heart.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; M15527; AAA42140.1; -;
 CC PIR; A29760; A29760.
 CC SQ SEQUENCE 106 AA; 11792 MW; F9EF8FDBFEABE3E CRC64;
 Query Match 32.5%; Score 50; DB 1; Length 106;
 Best Local Similarity 47.1%; Pred. No. 5.3;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 Oy 5 AGAARPRASWCALALL 21
 Db 2 AAHVRKSNWCQTYLL 18
 RESULT 7
 ID DPM3_HUMAN STANDARD; PRT; 92 AA.
 AC Q9P2X0; Q9BXN4; Q9BXN5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dolichol-phosphate mannosyltransferase subunit 3 (Dolichol-phosphate
 DE mannosyltransferase subunit 3) (Mannose-P-dolichol synthase subunit 3)
 DE

(DPM synthase subunit 3) (DPM synthase complex subunit 3) (Prostin 1).
 DPM3.
 Homo sapiens (Human).
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 C Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 C NCBI_TaxID=9606;
 P SEQUENCE FROM N.A. (ISOFORM 1), SEQUENCE OF 1-13, AND
 P CHARACTERIZATION.
 X MEDLINE=20296673; PubMed=1085346;
 T Maeda Y., Tanaka S., Hino J., Kangawa K., Kinoshita T.,
 T "Human dolichol-phosphate synthase consists of three subunits,
 T DPM1, DPM2 and DPM3."
 T EMBO J. 19:2475-2482 (2000).
 L SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 C MEDLINE=21313795; PubMed=11420690;
 C Manos E.J., Kim M.L., Kassisi J., Chang P.Y., Jones D.A.,
 C "Dolichol-phosphate-mannose-3 (DPM3)/prostin-1 is a novel
 C phospholipase C-gamma regulated gene negatively associated with
 C prostate tumor invasion."
 C Oncogene 20:2781-2790 (2001).
 C -1- FUNCTION: Stabilizer subunit of the dolichol-phosphate-mannose
 C synthase complex.
 C -1- SUBUNIT: Composed of three subunits; DPM1, DPM2 and DPM3.
 C Associated with DPM1 via its C-terminal domain and with DPM2 via
 C its N-terminal portion.
 C -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 C reticulum.
 C -1- ALTERNATIVE PRODUCTS:
 C Event=Alternative splicing; Named isoforms=2;
 C Name=1; Synonyms=Short;
 C IsoId=Q9P2X0-1; Sequence=Displayed;
 C Name=2; Synonyms=Long;
 C IsoId=Q9P2X0-2; Sequence=VSP_001308;
 C -1- SIMILARITY: Belongs to the DPM3 family.
 C
 C This SWISS-PROT entry is copyright. It is produced through a collaboration
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -
 C the European Bioinformatics Institute. There are no restrictions on its
 C use by non-profit institutions as long as its content is in no way
 C modified and this statement is not removed. Usage by and for commercial
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 C or send an email to license@isb-sib.ch).
 C
 C -----
 C EMBL: AB028128; BAA96291.1; -;
 C EMBL: AF312922; AAC28487.1; -;
 C EMBL: AF312923; AAC28486.1; -;
 C Genew: HGNC:3007; DPM3.
 C MIM: 605951; -;
 C TRANSMEM: 8
 C TRANSMEM: 37
 C TRANSFAC: 1
 C VARSPLIC: 1
 C
 C CONFLICT: 90 90
 C SEQUENCE 92 AA; 10080 MW; C350B3696842A877 CRC64;
 C
 C Query Match 32.1%; Score 49.5; DB 1; Length 92;
 C Best local Similarity 45.0%; Pred. No. 5.4;
 C Matches 9; Conservative 3; Mismatches 5; Indels 3; Gaps 1;
 Y 12 ASWCNALALL--WLAVERG 28
 b 5 AQLWGLALIGSTVALTTG 24
 EMBL 8
 LNS_HUMAN
 C CLN8_HUMAN STANDARD; PRT; 286 AA.
 C Q9UBX8; Q96195;
 T 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE CLN8 protein.
 GN CLN8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., VARIANT EPMR GUY-24, AND VARIANT ALA-155.
 RX MEDLINE=99438402; PubMed=10508524;
 RA Ranta S., Zhang Y., Ross B., Lonka L., Takunen E., Messer A.,
 RA Sharp J., Wheeler R., Kusumi K., Mole S., Liu W., Soares M.B.,
 RA Ronaldo M.F., Hirvasniemi A., de la Chapelle A., Gilliam T.C.,
 RA Lehesjoki A.-E.,
 RT "The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice
 RT are associated with mutations in CLN8."
 RL Nat. Genet. 23:233-236 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldi W.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshitsugu S., Carninci P., Prange C.,
 RA Raha S.S., Loguercio N.A., Peters G.J., Abramson R.D., Mallory S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,
 RA Fahy J., Hellon E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Maize M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP SUBCELLULAR LOCATION: AND MTTAGENESIS OF 283-LYS-LYS-284.
 RX MEDLINE=20320699; PubMed=10861296;
 RA Lonka L., Kytela A., Ranta S., Jalanko A., Lehesjoki A.-E.,
 RT "The neuronal ceroid lipofuscinoses CLN8 membrane protein is a
 RT resident of the endoplasmic reticulum."
 RL Hum. Mol. Genet. 9:1691-1697 (2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum and ER-Golgi intermediate compartment (ERGIC).
 CC -1- PTM: Does not seem to be N-glycosylated.
 CC -1- DISEASE: Defects in CLN8 are a cause of progressive epilepsy with
 CC mental retardation (EPMR) [MIM:600143]; also known as neuronal
 CC ceroid lipofuscinoses type 8 and Northern epilepsy. EPMR is an
 CC autosomal recessive disorder characterized by normal early
 CC development, onset of generalized seizures between 5 and 10 years,
 CC and subsequent progressive mental retardation. Biochemically, the
 CC disease is characterized by the intracellular accumulation of
 CC hydrophobic material, mainly ATP synthase subunit C.
 CC -1- SIMILARITY: Contains 1 TUC (TRAM/LAG1/CLN8) domain.
 CC -1- DATABASE: NAME=NCL CLN8;
 CC NOTE=Neural Ceroid Lipofuscinoses mutation db;
 CC WWW=<http://www.ucl.ac.uk/ncl/CLN8.html>.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----

```
DR EMBL, AF123757; AAF13115.1; -
DR EMBL, AF123758; AAF13116.1; -
DR EMBL, AF123759; AAF13117.1; -
DR EMBL, AF123760; AAF13118.1; -
DR EMBL, AF123761; AAF13119.1; -
DR EMBL, BC007725; AA077725.1; -
DR GenBank; HGNC:2079; CLN8.
DR MIM; 607837; -.
DR GO; GO:0005783; C:Endoplasmic reticulum; TAS.
DR GO; GO:0005793; C:ER-Golgi intermediate compartment; TAS.
DR GO; GO:0016021; C:Integral to membrane; TAS.
DR GO; GO:0007399; P:neurogenesis; TAS.
DR InterPro; IPR006634; TIG.
DR SMART; SM00724; TIG; 1.
DR PROSITE; PS50922; TIG; 1.
KW Transmembrane; Endoplasmic reticulum; Neuronal ceroid lipofuscinosis;
KW Disease mutation; Polymorphism; Epilepsy.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 62 84 POTENTIAL.
FT TRANSMEM 103 123 POTENTIAL.
FT TRANSMEM 131 151 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT DOMAIN 62 262 TIG.
FT SITE 283 286 ER-RETRIEVAL SIGNAL.
FT VARIANT 24 24 R->G (in EPMR).
FT VARIANT 155 155 /FTID=VAR_013174.
FT MUTAGEN 283 284 KK->RR: LOCALIZES TO THE GOLGI COMPLEX.
FT CONFLICT 225 225 N->S (IN REF. 2).
SQ SEQUENCE 286 AA; 32842 MW; 0BAEBCA516D1DC CRC64;

Query Match 31.8%; Score 49; DB 1; Length 286;
Best Local Similarity 58.3%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 AGAARGASMCW 16
DB 93 ADKARGQNCW 104

RESULT 9
PEXD CAEEL STANDARD; PRT; 330 AA.
AC 019951;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable peroxisomal membrane protein PEX13 (Peroxin-13).
GN P32A5.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodetinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Pauley A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN 12;
RP REVISIONS.
RA Waterston R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Component of the peroxisomal translocation machinery
CC with PEX14 and PEX17. Functions as a docking factor for the
CC predominantly cytoplasmic PIS1 receptor (PIS1/PEX5) (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal (By
CC similarity).
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U20864; AAC46665.2; -.
DR HSSP; P00519; IAB1.
DR WormPep; P32A5.6; CE28290.
DR InterPro; IPR007223; Peroxin-13_N.
DR InterPro; IPR001452; SH3.
DR Pfam; PF04088; Peroxin-13_N; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Peroxisome; Transport; Protein transport; Transmembrane; SH3 domain.
FT DOMAIN 1 192 LUMENAL (POTENTIAL).
FT TRANSMEM 193 213 POTENTIAL.
FT DOMAIN 214 330 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 235 300 SH3.
SQ SEQUENCE 330 AA; 35765 MW; 7DD647F07A0403 CRC64;

Query Match 31.8%; Score 49; DB 1; Length 330;
Best Local Similarity 43.5%; Pred. No. 19;
Matches 10; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

OY 6 GAARGASMCWALMLWAVPG 28
DB 187 GATRPASVWVPALFWVAIG 209

RESULT 10
INT_SYNY3 STANDARD; PRT; 519 AA.
AC P74055;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apolipoprotein N-acyltransferase (EC 2.3.1.-) (AIP N-acyltransferase).
GN LNT OR SLR0819.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN 11;
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yanada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: Transfers the fatty acyl group on membrane lipoproteins
CC (By similarity).
CC -1- PATHWAY: Lipoproteins biosynthesis.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the apolipoprotein N-acyltransferase
CC family.
CC -1- SIMILARITY: Contains 1 CN hydrolase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```


10 TISSUE SPECIFICITY: Ubiquitously expressed with high expression in
10 testis, moderate expression in the pancreas, spleen, prostate,
10 small intestine and colon. Very low expression is seen in brain
10 and skeletal muscle.
10
10 -1- SIMILARITY: Contains 3 MIR domains.
10
10 This SWISS-PROT entry is copyright. It is produced through a collaboration
10 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
10 the European Bioinformatics Institute. There are no restrictions on its
10 use by non-profit institutions as long as its content is in no way
10 modified and this statement is not removed. Usage by and for commercial
10 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
10 or send an email to license@isb-sib.ch).
10
10 EMBL; AB043007; BAB18277.1; -
10 EMBL; AF277316; AAK69113.1; -
10 EMBL; BC006248; AAH06248.1; -
10 PIR; JC7587; JC7587.
10 Genew; HGNC:10676; SDF2L1.
10 MIM; 607551; -
10 InterPro; IPR000886; ER_target_S.
10 InterPro; IPR003608; MIR.
10 Pfam; PF02815; MIR_3.
10 SMART; SMO0472; MIR_3.
10 PROSITE; PS00014; ER_TARGET; 1.
10 PROSITE; PS00919; MIR_3.
10 Endoplasmic reticulum; Signal; Repeat.
10 SIGNAL 1 28 POTENTIAL.
10 CHAIN 29 221 STROMAL CELL-DERIVED FACTOR 2-LIKE
10
10 DOMAIN 33 87 MIR 1.
10 DOMAIN 95 150 MIR 2.
10 SITE 151 205 MIR 3.
10 SITE 218 221 PREVENT SECRETION FROM ER (POTENTIAL).
10 CONFLICT 94 94 C->R (IN REF. 2 AND 3).
10 CONFLICT 162 162 L->F (IN REF. 2 AND 3).
10 SEQUENCE 221 AA; 23511 MW; 23EBBE14BE9E92BF CRC64;
10
10 Query Match 31.2%; Score 48; DB 1; Length 221;
10 Best Local Similarity 50.0%; Pred. No. 18;
10 Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
10
10 7 AARGASGWCALALMLAVVG 28
10 4 AGRGGAAPVLLGLLALVVG 25
10
10 RESULT 13
10 TCT AGRTS STANDARD; PRT; 233 AA.
10
10 P29112;
10 01-DEC-1992 (Rel. 24, Created)
10 01-DEC-1992 (Rel. 24, Last sequence update)
10 28-FEB-2003 (Rel. 41, Last annotation update)
10 24.9 kDa protein in p1CA locus (ORF1).
10 ATU3128 OR AGR_L 3363.
10 Agrobacterium tumefaciens (strain C58 / ATCC 33970), and
10 Agrobacterium tumefaciens.
10 Bacteria: Proteobacteria; Alphaproteobacteria; Rhizobiales;
10 Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
10 NCBI_TaxID=176299, 358;
10
10 SEQUENCE FROM N.A.
10 STRAIN=C58 / ATCC 33970;
10 MEDLINE=21608550; PubMed=11743193;
10
10 Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
10 Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
10 Chen Y., Paulsen I.T., Bisen J.A., Karp P.D., Bovee D., Str.,
10 Chapman P., Clendenning J., Decherage G., Gillet M., Grant C.,
10 Kutyavin T., Levy R., Li M.-D., McClelland E., Palmieri A.,
10 Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
10 Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
10 Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,

10 Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
10 Neeter B.W.;
10 "The genome of the natural genetic engineer Agrobacterium tumefaciens
10 C58".
10 Science 294:2317-2323(2001).
10
10 [2]
10 SEQUENCE FROM N.A.
10 STRAIN=C58 / ATCC 33970;
10 MEDLINE=21608551; PubMed=11743194;
10 Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
10 Qurollo B., Goldman B.S., Gao Y., Akenazi M., Halling C., Mullin L.,
10 Houmli K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu P.,
10 Wolan C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
10 Flanagan C., Crowell C., Garson J., Lomo C., Sear C., Strub G.,
10 Cielo C., Slater S.,
10 "Genome sequence of the plant pathogen and biotechnology agent
10 Agrobacterium tumefaciens C58".
10 Science 294:2323-2328(2001).
10
10 [3]
10 SEQUENCE FROM N.A.
10 STRAIN=A136;
10 MEDLINE=91317730; PubMed=1860822;
10 Rong L., Karcher S.-J., Gelvin S.B.,
10 "Genetic and molecular analyses of p1CA, a plant-inducible locus on
10 the Agrobacterium tumefaciens chromosome".
10 J. Bacteriol. 173:5110-5120(1991).
10
10 -1- FUNCTION: SEEM TO REGULATE THE SURFACE PROPERTIES OF THE
10 BACTERIUM IN THE PRESENCE OF PLANT CELLS OR PLANT CELL EXTRACTS.
10 MUTATIONS IN THIS PROTEIN ARE RESPONSIBLE FOR AN INCREASED
10 AGGREGATION OF THE BACTERIA IN THE PRESENCE OF PEA ROOT CAP
10 CELLS.
10
10 -1- INDUCTION: By certain acidic polysaccharides found in carrot root
10 extract. This induction may be regulated by the polygalacturonase.
10
10 This SWISS-PROT entry is copyright. It is produced through a collaboration
10 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
10 the European Bioinformatics Institute. There are no restrictions on its
10 use by non-profit institutions as long as its content is in no way
10 modified and this statement is not removed. Usage by and for commercial
10 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
10 or send an email to license@isb-sib.ch).
10
10 EMBL; AE009243; AAL43944.1; ALT_INIT.
10 EMBL; AE008370; AAK90258.1; ALT_INIT.
10 EMBL; M62814; AAA22103.1; -
10 PIR; B40364; B40364.
10 Complete proteome.
10 SEQUENCE 233 AA; 24935 MW; 4ED00E8C180195F7 CRC64;
10
10 Query Match 31.2%; Score 48; DB 1; Length 233;
10 Best Local Similarity 41.7%; Pred. No. 19;
10 Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
10
10 2 RRGAGARGASGWCALALMLAV 25
10 50 QRWANPASGKSPAIWARAVGLAM 73
10
10 RESULT 14
10 NID0 MOUSE STANDARD; PRT; 1245 AA.
10
10 P10493;
10 01-APR-1990 (Rel. 14, Created)
10 01-APR-1990 (Rel. 14, Last sequence update)
10 15-MAR-2004 (Rel. 43, Last annotation update)
10 Nidogen precursor (Enactin).
10 NID OR NID1 OR ENT.
10 Mus musculus (Mouse).
10 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
10 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
10 NCBI_TaxID=10090;
10
10 SEQUENCE FROM N.A., AND SEQUENCE OF 29-40.

MEDLINE=69079780; PubMed=3264556;
Duckin M.E.; Chakravarti S.; Bartos B.B.; Liu S.H.; Friedman R.L.;
Chung A.E.;
"Amino acid sequence and domain structure of entactin. Homology with
epidermal growth factor precursor and low density lipoprotein
receptor";
J. Cell Biol. 107:2749-2756(1988).
[2]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
MEDLINE=89231638; PubMed=2496973;
Mann K., Deutzmann R., Annalley M., Timpl R., Raimondi L.,
Yamada Y., Pan T.-C., Conway D., Chu M.-L.;
"Amino acid sequence of mouse nidogen, a multidomain basement
membrane protein with binding activity for laminin, collagen IV and
cells";
EMBO J. 8:65-72(1989).
[3]
SEQUENCE OF 1-251 FROM N.A.
STRAIN=BALB/c; TISSUE=Liver;
MEDLINE=94040771; PubMed=8224873;
Duckin M.E., Liu S.H., Reing J., Chung A.E.;
"Characterization of the 5' end of the mouse Ent gene encoding the
basement membrane protein, entactin.";
Gene 132:261-266(1993).
[4]
SEQUENCE OF 1207-1245 FROM N.A.
STRAIN=C57BL/6J X CBA/J;
MEDLINE=95324912; PubMed=7601446;
Duckin M.E., Wewer U.M., Chung A.E.;
"Bson organization of the mouse entactin gene corresponds to the
structural domains of the polypeptide and has regional homology to
the low-density lipoprotein receptor gene";
Genomics 26:219-228(1995).
[5]
PARTIAL SEQUENCE.
MEDLINE=86192477; PubMed=3084254;
Paulsson M., Deutzmann R., Dziadek M., Nowack H., Timpl R., Weber S.,
Engel J.;
"Purification and structural characterization of intact and
fragmented nidogen obtained from a tumor basement membrane";
Eur. J. Biochem. 156:467-478(1986).
[6]
CARBOHYDRATE-LINKAGE SITES, AND PARTIAL SEQUENCE.
MEDLINE=93316903; PubMed=8326911;
Fujiiwara S., Shinkai H., Mann K., Timpl R.;
"Structure and localization of O- and N-linked oligosaccharide chains
on basement membrane protein nidogen.";
Matrix 13:215-222(1993).
[7]
INTERACTION WITH FBLN1.
MEDLINE=97446166; PubMed=9299350;
Adam S., Goehring W., Wiedemann H., Chu M.-L., Timpl R., Kostka G.;
"binding of fibulin-1 to nidogen depends on its C-terminal globular
domain and a specific array of calcium-binding epidermal growth
factor-like (EGF) modules";
J. Mol. Biol. 272:226-236(1997).
[8]
INTERACTION WITH FBLN1.
MEDLINE=21474010; PubMed=11589703;
Ries A., Goehring W., Fox J.W., Timpl R., Sasaki T.;
"Recombinant domains of mouse nidogen-1 and their binding to basement
membrane proteins and monoclonal antibodies";
Eur. J. Biochem. 268:5119-5128(2001).
[9]
FUNCTION: Sulfated glycoprotein which is widely distributed in
basement membranes and that is tightly associated with laminin.
Also binds to collagen IV. It probably has a role in cell-
extracellular matrix interactions.
[10]
SUBUNIT: Interacts with FBLN1.
[11]
SUBCELLULAR LOCATION: Basement membranes.
[12]
PTM: N- and O-glycosylated.
[13]
SIMILARITY: Contains 6 EGF-like domains.
[14]
SIMILARITY: Contains 1 thyroglobulin type-I domain.
[15]
SIMILARITY: Contains 5 IRII-receptor WYTD domains.

[illegible]


```

T DISULFID 787 798 BY SIMILARITY.
T DISULFID 804 815 BY SIMILARITY.
T DISULFID 809 824 BY SIMILARITY.
T DISULFID 826 837 BY SIMILARITY.
T DISULFID 1210 1221 BY SIMILARITY.
T DISULFID 1217 1230 BY SIMILARITY.
T DISULFID 1232 1241 BY SIMILARITY.
T SITE 700 702 CELL ATTACHMENT SITE.
T CARBOHYD 187 187 N-LINKED (GLCNAC. . .).
T CARBOHYD 299 299 O-LINKED.
T CARBOHYD 331 331 O-LINKED.
T CARBOHYD 337 337 O-LINKED.
T CARBOHYD 345 345 O-LINKED.
T CARBOHYD 348 348 O-LINKED (PARTIAL. . .).
T CARBOHYD 415 415 N-LINKED (GLCNAC. . .).
T CARBOHYD 920 920 O-LINKED.
T CARBOHYD 933 933 O-LINKED.
T CONFLICT 170 170 P -> L (IN REF. 2).
T CONFLICT 659 659 R -> K (IN REF. 2).
T CONFLICT 967 967 R -> A (IN REF. 2).
Q SEQUENCE 1245 AA; 136622 MW; 8FBE276F29BE6D2 CRC64;

```

```

Query Match 31.2%; Score 48; DB 1; Length 1245;
Best Local Similarity 38.8%; Pred. No. 82;
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

Y 12 ASWCWALALMLAVPG 28
b 10 AMMTWALLQLLVGPG 26

```

```

RESULT 15
JGB_ECOLI

```

```

D YJGB_ECOLI STANDARD; PRT; 339 AA.
P27250; P76812;

```

```

T 01-AUG-1992 (Rel. 23, Created)
T 01-FEB-1995 (Rel. 31, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E Hypothetical zinc-type alcohol dehydrogenase-like protein yjgb.
N YJGB OR B4269.
S Escherichia coli.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Escherichia.
C NCBI_TaxID=562;

```

```

P SEQUENCE FROM N.A.
C STRAIN=B.

```

```

A Pucci M.J., Discotto L.F., Dougherty T.J.;
N Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
P [2]

```

```

P SEQUENCE FROM N.A.
C STRAIN=K12 / MG1655;

```

```

X MEDLINE=95334362; PubMed=7610040;
A Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
N Blatcher F.R.;
T "Analysis of the Escherichia coli genome VI: DNA sequence of the
L region from 92.8 through 100 minutes."
C Nucleic Acids Res. 23:2105-2119(1995).
C -1- COPACTOR: Binds 2 zinc ions per subunit (By similarity).
C -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
C family.

```

```

C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb.ch/announce/
C or send an email to license@isb-sib.ch).

```

```

R EMBL; M96355; AAA72122.1; -
R EMBL; U14003; AAA97166.1; ALT_INIT.
R EMBL; AE000497; AAC77226.1; ALT_INIT.

```

```

DR EcoGene; EGI1436; Yjgb.
DR InterPro; IPR002128; Adh_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; Adh_zinc_N; 1.
DR PROSITE; PS00059; Adh_ZINC; 1.
KW Hypothetical protein; Oxidoreductase; Zinc; Metal-binding;
KW Complete proteome.
FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 63 63 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 96 96 ZINC 2 (BY SIMILARITY).
FT METAL 99 99 ZINC 2 (BY SIMILARITY).
FT METAL 102 102 ZINC 2 (BY SIMILARITY).
FT METAL 110 110 ZINC 2 (BY SIMILARITY).
FT METAL 152 152 ZINC 2 (BY SIMILARITY).
FT CONFLICT 331 339 YRVLRADF -> TAWC (IN REF. 1).
SQ SEQUENCE 339 AA; 36502 MW; 0854DDEF16B9EE CRC64;

```

```

Query Match 30.5%; Score 47; DB 1; Length 339;
Best Local Similarity 45.0%; Pred. No. 36;
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 1 MRRGAGARGRASWCWALAL 20
Db 119 MNRGCFARLRADWDQVPIPL 138

```

```

Search completed: June 8, 2004, 14:04:38
Job time : 3.16667 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: June 8, 2004, 13:59:08 ; Search time 5.17222 Seconds
(without alignments)
1708.066 Million cell updates/sec

file: US-10-010-050A-2_COPY_1_28
effect score: 154
sequence: 1 MRRGGAARGRASCMALALMLAVPG 28

scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

set-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPRENBL_25.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_recent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66	42.9	606	17	Q9HMT4
2	59	38.3	412	16	Q92VB2
3	59	38.3	588	16	Q82G94
4	58	37.7	146	11	Q8CEK6
5	57.5	37.3	343	16	Q8G419
6	57	37.0	366	10	Q84X72
7	56	36.4	214	16	Q69464
8	56	36.4	485	16	Q88SV3
9	55.5	36.0	412	10	Q8XW40
10	55.5	36.0	695	16	Q8XW40
11	55	35.7	147	10	Q7X917
12	55	35.7	304	16	Q7WFK8
13	55	35.7	707	10	Q7W451
14	55	35.7	707	10	Q92R26
15	55	35.7	3019	12	Q92529
16	54.5	35.4	205	16	Q8PDR8

17	54.5	35.4	435	10	Q9XE69	Q9XE69 sorghum bic
18	54.5	35.4	836	16	Q89HK3	Q89HK3 bradyrhizob
19	54	35.1	214	10	Q9XE06	Q9XE06 sorghum bic
20	54	35.1	232	16	Q9XE04	Q9XE04 xyella fas
21	54	35.1	232	16	Q87EH4	Q87EH4 xyella fas
22	54	35.1	266	2	Q8L311	Q8L311 vitreoscill
23	53.5	34.7	126	2	Q8W997	Q8W997 alcaligenes
24	53	34.4	83	9	Q7Y4J5	Q7Y4J5 streptococ
25	53	34.4	160	16	Q82HW2	Q82HW2 streptococ
26	53	34.4	224	2	Q9EY24	Q9EY24 xanthomonas
27	53	34.4	428	16	Q8UD14	Q8UD14 agrobacteri
28	52.5	34.1	201	16	Q8PE05	Q8PE05 xanthomonas
29	52.5	34.1	333	16	Q8NRA3	Q8NRA3 corynebacte
30	52.5	34.1	334	2	Q9KML7	Q9KML7 corynebacte
31	52	33.8	119	10	Q7XT22	Q7XT22 oryza sativ
32	52	33.8	455	16	Q88D16	Q88D16 pseudomonas
33	52	33.8	658	4	Q9BX78	Q9BX78 homo sapien
34	52	33.8	667	4	Q9BX79	Q9BX79 homo sapien
35	52	33.8	667	4	Q8TB21	Q8TB21 homo sapien
36	52	33.8	708	4	Q7Z3U9	Q7Z3U9 homo sapien
37	51.5	33.4	543	10	Q94DP0	Q94DP0 oryza sativ
38	51.5	33.4	1146	13	Q90584	Q90584 gallus gall
39	51.5	33.4	1252	10	Q7XU28	Q7XU28 oryza sativ
40	51.5	33.4	1252	10	Q8GU77	Q8GU77 oryza sativ
41	51	33.1	102	16	Q7UB52	Q7UB52 shigella fl
42	51	33.1	248	16	Q8PF85	Q8PF85 xanthomonas
43	51	33.1	410	16	Q9ZTU9	Q9ZTU9 rhizobium m
44	51	33.1	456	16	Q93JK1	Q93JK1 streptococ
45	51	33.1	463	10	Q8H053	Q8H053 oryza sativ

ALIGNMENTS

RESULT 1	ID	PRELIMINARY:	PRT:	606 AA.
Q9HMT4	Q9HMT4	Q9HMT4		
AC	Q9HMT4	Q9HMT4		
AD	Q9HMT4	Q9HMT4		
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 24, Last annotation update)			
DE	Sulfate transport system permease protein.			
GN	CYS11 OR WAG2396G.			
OS	Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).			
OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;			
OC	Halobacteriaceae; Halobacterium.			
OX	NCBI_TaxID=64091;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20504483; PubMed=11016950;			
RA	Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,			
RA	Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna U.,			
RA	Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,			
RA	Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,			
RA	Maddocks D.G., Jablonski P.E., Krebs M.P., Angwine C.M., Dale H.,			
RA	Isehnacker T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,			
RA	Alan M., Freitas T., Hou S., Daniels C.T., Dennis P.P., Omer A.D.,			
RA	Bhardt H., Lowe T.W., Liang P., Riley M., Hood L., Dassarma S.;			
RT	"Genome sequence of Halobacterium species NRC-1".			
RT	Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).			
-1-	FUNCTION: PROBABLY PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT			
-1-	SYSTEM. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE			
CC	SUBSTRATE ACROSS THE MEMBRANE (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-			
CC	-1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-			
CC	PROTEIN-DEPENDENT TRANSPORT SYSTEMS.			
DR	EMBL: AE005121; AAC20487.1; -			
DR	PIR: C84390; C84390.			
DR	GO: GO:0016021; C:integral to membrane; IEA.			
DR	GO: GO:0003700; F:transcription factor activity; IEA.			
DR	GO: GO:0005215; F:transporter activity; IEA.			
DR	GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.			
DR	GO: GO:0006810; P:transport; IEA.			

DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR000847; HTH_LYSR.
DR Pfam; PF000529; BPD transp. 2.
DR PROSITE; PS00402; BPD_TRANS_INN_MEMBER; 1.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
DR Transmembrane; Transp; Complete proteome.
KW Transmembrane; Transp; Complete proteome.
SQ SEQUENCE 606 AA; 63066 MW; 2E64B9A20298A28C CRC64;

Query Match 42.9%; Score 66; DB 17; Length 606;
Best Local Similarity 53.8%; Pred. No. 1.1;
Matches 14; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 3 RGAGARGASGCMALALMLAVPG 28
DB 508 RALGASRGALMDVPLVMEGVAG 533

RESULT 2
Q92VB2 PRELIMINARY; PRT; 412 AA.
AC Q92VB2:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE Hypothetical membrane protein SMD21292.
GN RB0796 OR SMD21292.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie J.,
RA Golding B., Pehler A.;
RT "The complete sequence of the 1,683-Rb pSymb megaplasmid from the N2-
fixing endosymbiont Sinorhizobium meliloti";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).
DR EMBL; AL603644; CAC49196.1; -.
DR PIR; D95941; D95941.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000345; CytoC_heme_BS.
DR PROSITE; PS00190; CYTOCHROME_C7_1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 412 AA; 46126 MW; 8088B8E0C66F8838 CRC64;

Query Match 38.3%; Score 59; DB 16; Length 412;
Best Local Similarity 37.0%; Pred. No. 6.9;
Matches 10; Conservative 6; Mismatches 5; Indels 6; Gaps 1;

QY 7 AARGRASGCMW-----DALMLAVP 27
DB 276 ANKGRPTWTWLTATVLTALIMLSTVP 302

RESULT 3
Q82G94 PRELIMINARY; PRT; 588 AA.
AC Q82G94:
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DE Hypothetical protein.
GN SAV4004.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
NCBI_TaxID=33903;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa U., Hanamoto A., Takahashi C.,
RA Shinoue M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa U., Hanamoto A., Shinoue M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis";
RL Nat. Biotechnol. 21:526-531 (2003).
DR EMBL; AP005037; BAC71716.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 588 AA; 63428 MW; D8DC5F23BE49A321 CRC64;

Query Match 38.3%; Score 59; DB 16; Length 588;
Best Local Similarity 29.5%; Pred. No. 9.5;
Matches 13; Conservative 5; Mismatches 4; Indels 22; Gaps 1;

QY 3 RGAGARGASGCMALALMLAVP 24
DB 167 RGAGSRTKLYATKNTLIGVSGRYGDKSSWAGLWLA 210

RESULT 4
Q8CEK6 PRELIMINARY; PRT; 146 AA.
ID Q8CEK6:
AC Q8CEK6:
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein (Fragment).
GN 9330132005R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
DR EMBL; AK020369; BAC25627.1; -.
DR MED; MGI:1924802; 9330132005R1K.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 146 AA; 16423 MW; AF4B10EB630CB6D5 CRC64;

Query Match 37.7%; Score 58; DB 11; Length 146;
Best Local Similarity 52.4%; Pred. No. 3.7;
Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 2 RRGAGARGASGCMALALIM 22
DB 44 RRGAPRRSQPSSCWCGLIM 64

RESULT 5
Q8G419 PRELIMINARY; PRT; 343 AA.
ID Q8G419

086419:
01-MAR-2003 (TREMBlrel. 23, Created)
01-MAR-2003 (TREMBlrel. 23, last sequence update)
01-MAR-2003 (TREMBlrel. 23, last annotation update)
Hypothetical protein.
B1576.
Bifidobacterium longum.
Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
Bifidobacteriaceae; Bifidobacterium.
NCBI_TaxID=216816;
[1]
SEQUENCE FROM N.A.
STRAIN=NCC 2705;
MEDLINE=22294977; PubMed=12381787;
Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
Pessi G., Zwalen M.-C., Desiere F., Bork P., Delley M.,
Pridmore R.D., Arigoni F.,
"The genome sequence of Bifidobacterium longum reflects its adaptation
to the human gastrointestinal tract."
Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
EMBL; AB014793; AAN25367.1; -
Hypothetical protein; Complete proteome.
SEQUENCE 343 AA, 37507 MW, BAB35FF32D22E059 CRC64;

Query Match 37.3%; Score 57.5; DB 16; Length 343;
Best Local Similarity 54.2%; Pred. No. 9.3;
Matches 13; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

9 RGR-----ASWCWALALMLAVP 27
|||:|||||:
266 RGRCSYVAASWVWVALLMLAVP 289

SUITE 6
4X72

PRELIMINARY; PRT; 366 AA.

084X72:
01-JUN-2003 (TREMBlrel. 24, Created)
01-JUN-2003 (TREMBlrel. 24, last sequence update)
01-JUN-2003 (TREMBlrel. 24, last annotation update)
CR066 protein.

Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
NCBI_TaxID=3055;
[1]

SEQUENCE FROM N.A.
Li J.B., Lin S., Jia H., Wu H., Roe B.A., Kulp D., Stormo G.D.,
Dutcher S.K.,
"Finished genomic sequence in Chlamydomonas."
Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.

EMBL; AY207499; AAO32622.1; -
SEQUENCE 366 AA; 38816 MW; 99F528BB9E625BB CRC64;

Query Match 37.0%; Score 57; DB 10; Length 366;
Best Local Similarity 45.5%; Pred. No. 11;
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

4 GAGAGRAGASGCMALALMLAV 25
|||:|||||:
315 GGGGQPGRRRCWAEGLMVEV 336

SUITE 7
9464

PRELIMINARY; PRT; 214 AA.

069464:
01-AUG-1998 (TREMBlrel. 07, Created)
01-AUG-1998 (TREMBlrel. 07, last sequence update)
01-JUN-2003 (TREMBlrel. 24, last annotation update)
Putative integral membrane protein (conserved integral membrane
protein).
ML1666 OR MLCB1243.07.

OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=11769;
[1]

RP SEQUENCE FROM N.A.

STRAIN=TN;

RC MEDLINE=21128732; PubMed=11234002.

RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,

Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,

Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,

RA Holroyd S., Hornsby T., Jagels K., Jacrot C., Maclean J., Moule S.,

RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,

RA Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,

RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,

RA Barrell B.G.,
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).

DR EMBL; AL023635; CAI9188.1; -
DR EMBL; AL583923; CAC30619.1; -
DR PIR; T44701; T44701.
DR Leproma; ML1666; -
KM Complete proteome.

SEQUENCE 214 AA; 23402 MW; 5B48414709828FA0 CRC64;

Query Match 36.4%; Score 56; DB 16; Length 214;
Best Local Similarity 41.7%; Pred. No. 9.5;
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

3 RGAGARAGASGCMALALMLAV 26
|||:|||||:
175 RTSARAGIYQRMSTATLWFTV 198

RESULT 8
098SV3

PRELIMINARY; PRT; 485 AA.

01-JUN-2003 (TREMBlrel. 24, Created)
01-JUN-2003 (TREMBlrel. 24, last sequence update)
01-OCT-2003 (TREMBlrel. 25, last annotation update)
DE Cardiolipin synthetase 2 (EC 2.7.8.-).
GN C15 OR IP 3273.

OS Lactobacillus plantarum.

OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.

OX NCBI_TaxID=1590;
[1]

RP SEQUENCE FROM N.A.

STRAIN=NCTMB 8826 / WCFS1;
MEDLINE=22480296; PubMed=12566566;

RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,

RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,

RA Fiers M.W.E.J., Stiekema M., Klein Lankhorst R.M., Bron P.A.,

RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,

RA De Vos W.M., Siezen R.O.,
RT "Complete genome sequence of Lactobacillus plantarum WCFS1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).

DR EMBL; AL935261; CAD65406.1; -
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.

DR InterPro; IPR001736; PFD.

DR Pfam; PF00614; PUDc; 2.

DR PROSITE; PS50035; PUD; 2.

DR TRANSFERASE; Complete proteome.

SEQUENCE 485 AA; 55778 MW; C725C28841DF8D31 CRC64;

Query Match 36.4%; Score 56; DB 16; Length 485;
Best Local Similarity 46.7%; Pred. No. 20;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

12 ASWCWALALMLAVP 26

36 ATWAMLLVLTWLPRT 50

RESULT 9

12XSYO PRELIMINARY; PRT; 412 AA.

01-OCT-2003 (TREMBlrel. 25, Created)

01-OCT-2003 (TREMBlrel. 25, Last sequence update)

01-OCT-2003 (TREMBlrel. 25, Last annotation update)

OSUNB0056F09.16 protein.

OSUNB0056F09.16.

Oryza sativa (Rice).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

NCBI_TaxID=4530;

SEQUENCE FROM N.A.

Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,

Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,

Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,

Shang H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,

Huang Y.C., Li Y., Zhu J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,

Chen L., Fan D.L., Wang Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,

Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,

Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,

Zhang R.Q., Guan J.P., Hong G.F.,

Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

EMBL; AL606602; CAE01753.1; -

SEQUENCE 412 AA; 45803 MW; 1A8FE532333606C8 CRC64;

Query Match 36.0%; Score 55.5; DB 10; Length 412;

Best Local Similarity 40.7%; Pred. No. 20;

Matches 11; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

4 GAGAGAG-----RASWCALALMLAV 25

19 GSGGSGGLPRATRKEMWMSGILTKAV 45

RESULT 10

8XWMA PRELIMINARY; PRT; 695 AA.

01-MAR-2002 (TREMBlrel. 20, Created)

01-MAR-2002 (TREMBlrel. 20, Last sequence update)

01-OCT-2003 (TREMBlrel. 25, Last annotation update)

Probable bifunctional: uroporphyrin-III C-methyltransferase and

uroporphyrinogen-III synthase transmembrane protein

(EC 2.1.1.107).

RSC2356 OR RS01190.

Ralstonia solanacearum (Pseudomonas solanacearum).

Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

Burkholderiaceae; Ralstonia.

NCBI_TaxID=305;

SEQUENCE FROM N.A.

STRAIN=GMT1000.

MEDLINE=21681879; PubMed=11923852;

Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,

Arlat M., Billault A., Brotilier P., Camus J.C., Cactolico L.,

Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,

Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,

Stignier P., Thebaud P., Whalen M., Winkler P., Levy M.,

Weisenbach J., Boucher C.A.;

"Genome sequence of the plant pathogen Ralstonia solanacearum";

Nature 415:497-502(2002).

EMBL; AL646068; CAD16063.1; -

GO; GO:0008169; F:methyltransferase activity; IEA.

GO; GO:0016740; F:transferase activity; IEA.

GO; GO:0004851; F:uroporphyrin-III C-methyltransferase activity; IEA.

DR GO:0004852; F:uroporphyrinogen-III synthase activity; IEA.

DR GO:0006783; P:heme biosynthesis; IEA.

DR InterPro; IPR007470; DUF513.

DR InterPro; IPR003754; HEM4_synth.

DR Pfam; PF02602; HEM4; 1.

DR Pfam; PF04375; HemX; 1.

KM Transferase; Methyltransferase; Complete proteome.

SEQUENCE 695 AA; 73670 MW; 1649C5BA781DF05 CRC64;

Query Match 36.0%; Score 55.5; DB 16; Length 695;

Best Local Similarity 52.0%; Pred. No. 32;

Matches 13; Conservative 1; Mismatches 10; Indels 1; Gaps 1;

5 AGAAGRASM-CALALLMLAVPG 28

344 AGAGRGGLWMLALMLAVVAG 368

RESULT 11

07X917 PRELIMINARY; PRT; 147 AA.

01-OCT-2003 (TREMBlrel. 25, Created)

01-OCT-2003 (TREMBlrel. 25, Last sequence update)

01-OCT-2003 (TREMBlrel. 25, Last annotation update)

OSUNB0040D17.2 protein (OSUNB0033G08.19 protein).

OSUNB0040D17.2 OR OSUNB0033G08.19.

Oryza sativa (Rice).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

NCBI_TaxID=4530;

SEQUENCE FROM N.A.

Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,

Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Wang Q.J., Zhang L.,

Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,

Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,

Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,

Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,

Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,

Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,

Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;

Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

EMBL; AL606989; CAE02003.1; -

SEQUENCE 147 AA; 14767 MW; 024314375754B785 CRC64;

Query Match 35.7%; Score 55; DB 10; Length 147;

Best Local Similarity 33.3%; Pred. No. 9.2;

Matches 9; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

2 FRGAGAGRASWCALMLAVVPG 28

99 FRACSSRRDGMWMSCVMAQITPG 125

RESULT 12

Q7WEK8 PRELIMINARY; PRT; 304 AA.

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

Q7WEK8

PROTEIN C AND MRNA (BY SIMILARITY).

EMBL: D84252; BAA32664.1; -.
HSSP: E27958; IAIIV.
MEROPS; U39.001; -.
GO: GO:0016021; C:integral to membrane; IEA.
GO: GO:0019028; C:viral capsid; IEA.
GO: GO:0019031; C:viral envelope; IEA.
GO: GO:0005524; F:ATP binding; IEA.
GO: GO:0008026; F:ATP dependent helicase activity; IEA.
GO: GO:0003723; F:RNA binding; IEA.
GO: GO:0009368; F:RNA-directed RNA polymerase activity; IEA.
GO: GO:0009236; F:serine-type peptidase activity; IEA.
GO: GO:0005198; F:structural molecule activity; IEA.
GO: GO:0016740; F:transferase activity; IEA.
GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
GO: GO:0006350; P:transcription; IEA.
GO: GO:0019079; P:viral genome replication; IEA.
GO: GO:0019087; P:viral transformation; IEA.
InterPro: IPR009003; Cys_Ser_trypsin.
InterPro: IPR001410; DEAD.
InterPro: IPR002522; HCV_capsid.
InterPro: IPR002521; HCV_core.
InterPro: IPR002519; HCV_env.
InterPro: IPR002531; HCV_NS1.
InterPro: IPR002518; HCV_NS2.
InterPro: IPR000745; HCV_NS4.
InterPro: IPR001490; HCV_NS4B.
InterPro: IPR002868; HCV_NS5A.
InterPro: IPR002166; HCV_NS5A.
InterPro: IPR001650; Helicase_C.
InterPro: IPR003006; IG_MHC.
InterPro: IPR004109; peptidase_C9.
InterPro: IPR007095; RNA_pol_DS_PS.
InterPro: IPR007094; RNA_pol_PSVir.
Pfam: PF01543; HCV_capsid; 1.
Pfam: PF01542; HCV_core; 1.
Pfam: PF01539; HCV_env; 1.
Pfam: PF01560; HCV_NS1; 1.
Pfam: PF01538; HCV_NS2; 1.
Pfam: PF02907; HCV_NS3; 1.
Pfam: PF01006; HCV_NS4a; 1.
Pfam: PF01001; HCV_NS4b; 1.
Pfam: PF01506; HCV_NS5a; 1.
Pfam: PF00271; helicase_C; 1.
Pfam: PF00998; Viral_RdRp; 1.
ProDom: PD186062; HCV_NS1; 1.
SMART; SM00487; DEXDC; 1.
PROSITE; PS00290; IG_MHC; 1.
Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.

SEQUENCE 3019 AA; 328230 MW; E26750E07BCBC310 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

un on: June 8, 2004, 13:55:43 ; Search time 7.95833 Seconds

(without alignments)
1065.101 Million cell updates/sec

file: US-10-010-050a-2_COPY_1_30

sequence score: 169
1 MRRGAGAAAGRAAAGCAALALMLAVVPGWS 30

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_29Jan04:*

1: geneseqp1980s:*\n2: geneseqp1990s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	169	100.0	347	2	AAV41323 Human sec
2	169	100.0	358	4	AAW93870 Human pol
3	162	95.9	346	2	AAW92967 Human zst
4	60.5	35.8	192	4	AAU29418 Human G p
5	60.5	35.8	192	5	ABG60706 Novel G p
6	59.5	35.2	127	4	AAU14162 Human nov
7	58	34.3	73	4	ABG29252 Novel hum
8	58	34.3	188	5	ABU60973 Lung spec
9	57.5	34.0	19938	6	ABP76682 Streptomy
10	56	33.1	214	5	ABU05547 M. tuberc
11	56	33.1	403	5	ABU65190 Human NOV
12	56	33.1	1093	5	AAW41001 Human myo
13	55.5	32.8	371	4	AAW93279 Human pro
14	55.5	32.8	371	5	AAO18510 Human ins
15	55.5	32.8	371	5	ABG32373 Human hns
16	55.5	32.8	421	6	ADA10702 Human PHB
17	54.5	32.2	169	2	AAV59780 Human ntr
18	54.5	32.2	169	2	AAV60506 Human ntr
19	54	32.0	78	4	ABU16658 Human ner
20	53.5	31.7	95	6	ABR58357 Bcl1040 p
21	53.5	31.7	117	4	AAU54332 Propionib
22	53.5	31.7	117	6	ABW50851 Propionib
23	53	31.4	34	2	AAV43189 Thiredox
24	53	31.4	65	2	AAU00765 Ption pro
25	53	31.4	117	6	ADA54650 Human pro

26	53	31.4	211	4	ABG01709	Abg01709 Novel hum
27	53	31.4	243	4	AAU46489	Aau46489 Propionib
28	53	31.4	243	6	ABW43008	Abw43008 Propionib
29	53	31.4	346	5	ABG96249	Abg96249 Maize per
30	53	31.4	532	3	AAV84593	Aav84593 Amino aci
31	53	31.4	600	4	ABG07764	Abg07764 Novel hum
32	52.5	31.1	199	4	AAU28370	Aau28370 Novel hum
33	52.5	31.1	199	7	ADG09107	Adg09107 Novel pro
34	52.5	31.1	199	7	ADG09108	Adg09108 Novel pro
35	52.5	31.1	206	3	AAW42500	Aaw42500 Human ORF
36	52.5	31.1	244	4	AAW42195	Aaw42195 Human ORF
37	52.5	31.1	267	4	AAW79568	Aaw79568 Corynebac
38	52.5	31.1	333	4	AAW91017	Aaw91017 C glutami
39	52.5	31.1	341	3	AAV69880	Aav69880 B. lactof
40	52.5	31.1	360	6	ABU22472	Abu22472 Protein e
41	52.5	31.1	442	1	AAV90475	Aav90475 N-termina
42	52.5	31.1	551	2	AAW76503	Aaw76503 Cattle MI
43	52.5	31.1	575	1	AAV70195	Aav70195 Sequence
44	52.5	31.1	575	1	AAV90547	Aav90547 Bovine Mt
45	52.5	31.1	575	2	AAW76500	Aaw76500 Cattle MI

ALIGNMENTS

RESULT 1	AAV41323	standard; protein; 347 AA.
AC	AAV41323;	
DT	02-DEC-1999	(first entry)
DE	Human secreted protein encoded by gene 16 clone HMZAD77.	
XX	Human; secreted protein; fusion protein; gene therapy; protein therapy;	
KW	diagnosis; tissue; cancer; tumor; neurodegenerative disorder; leukaemia;	
KW	developmental abnormality; foetal deficiency; blood; allergy; renal;	
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;	
KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;	
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;	
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;	
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.	
OS	Homo sapiens.	
XX	W09947540-A1.	
XX	23-SEP-1999.	
PD	18-MAR-1999;	99WO-US005804.
XX	19-MAR-1998;	98US-0078563P.
XX	19-MAR-1998;	98US-0078566P.
XX	19-MAR-1998;	98US-0078573P.
XX	19-MAR-1998;	98US-0078574P.
XX	19-MAR-1998;	98US-0078577P.
XX	19-MAR-1998;	98US-0078578P.
XX	19-MAR-1998;	98US-0078579P.
XX	19-MAR-1998;	98US-0078581P.
XX	01-APR-1998;	98US-0080312P.
XX	01-APR-1998;	98US-0080313P.
XX	01-APR-1998;	98US-0080314P.
XX	(HUMA-) HUMAN GENOME SCT INC.	
XX	Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;	
XX	Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;	
XX	Shi Y, Moore PA,	
XX	WPI, 1999-562050/47.	
XX	N-PSDB; AAZ24826.	

XX New isolated human genes, useful for diagnosis and treatment of e.g.
XX cancers, neurological disorders, immune diseases, inflammation or blood
XX disorders.

XX Claim 11; Page 367-368; 484pp; English.

XX This sequence represents a secreted human protein encoded by the gene
XX clone detailed in the descriptor line. The gene can be used to generate
XX fusion proteins by linking to the gene to a human immunoglobulin Fc
XX portion (e.g. AA224802) for increasing the stability of the fused protein
XX as compared to the human protein only. The invention relates to 95 novel
XX genes and their fragments (nucleic acid sequences; AA224811-224907; amino
XX acid sequences AA41308-414104) which are useful for preventing, treating
XX or ameliorating medical conditions e.g. by protein or gene therapy. Also,
XX pathological conditions can be diagnosed by determining the amount of the
XX new polypeptides in a sample or by determining the presence of mutations
XX in the new polynucleotides. Specific uses are described for each of the
XX 95 polynucleotides, based on which tissues they are most highly expressed
XX in (see AA224811 for described uses)

XX Sequence 347 AA;

Query Match 100.0%; Score 169; DB 2; Length 347;
Best Local Similarity 100.0%; Pred. No. 8.7e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 MRRGAGAGRAGSCWALALMLAVVPGMS 30
b 1 MRRGAGAGRAGSCWALALMLAVVPGMS 30

RESULT 2
AM93870
D AAM93870 standard; protein; 358 AA.

C AAM93870;

T 06-NOV-2001 (first entry)

X Human polypeptide, SEQ ID NO: 3978.

W Human; full length cDNA; cDNA synthesis; oligo-capping.

S Homo sapiens.

X EP130094-A2.

D 05-SEP-2001.

F 07-JUL-2000; 2000EP-00114089.

R 08-JUL-1999; 99JP-00194486.

R 11-JAN-2000; 2000JP-00118774.

R 02-MAY-2000; 2000JP-00183765.

X (HELI-) HELIX RES INST.

I Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

I Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

R WPI; 2001-524255/58.

R N-PSDB; AAK94829.

S Claim 8; SEQ ID NO 3978; 1380pp + Sequence listing; English.

C The invention relates to primers for synthesising full length cDNA
C clones. 830 cDNA molecules encoding a human protein have been isolated
C and nucleotide sequences of 5' - and 3' -ends of the cDNA molecules have
C been determined. Primers for synthesising the full length cDNA are useful

CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a polypeptide encoded by a full length
CC human cDNA of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in CD-ROM
CC format directly from EPO

SQ Sequence 358 AA;

Query Match 100.0%; Score 169; DB 4; Length 358;
Best Local Similarity 100.0%; Pred. No. 9e-15;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRGAGAGRAGSCWALALMLAVVPGMS 30
Db 13 MRRGAGAGRAGSCWALALMLAVVPGMS 42

RESULT 3
AAM92967

ID AAM92967 standard; protein; 346 AA.

AC AAM92967;

DT 14-MAY-1999 (first entry)

DE Human zsig46 protein.

XX Secreted protein; zsig46; human; chromosome 13; thyroid; disease;

XX hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;

XX Hirschsprung's disease; neuronal ceroid-lipofucinosi; Wilson disease;

XX Reiger syndrome; immunoassay; detection; anti-idiotypic antibody;

XX therapy; diagnostic.

XX Homo sapiens.

XX MO9905275-A1.

XX 04-FEB-1999.

XX 24-JUL-1998; 98WO-US015431.

XX 24-JUL-1997; 97US-0053613P.

XX (ZYMO) ZYMOGENETICS INC.

XX Sheppard PO, Gilbertson DG;

XX WPI; 1999-142930/12.

XX N-PSDB; AAX02855.

XX New secreted polypeptide, zsig46, and its fragments, related fusion

XX proteins - used for diagnosis and treatment of thyroid disorders or

XX diseases involving genes on chromosome 13.

XX Claim 3; Page 90-91; 101pp; English.

CC This invention describes the isolation of a novel human secreted protein,
CC zsig46 encoded by a gene on chromosome 13 which is mainly expressed in
CC the thyroid. This product can be used to study secretion of proteins from
CC cells and also to treat or prevent deficient expression of zsig46, which
CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'
CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that
CC involve genes in the same region of chromosome 13 (e.g. Hirschsprung's
CC disease, neuronal ceroid-lipofucinosi, Wilson disease and Reiger
CC syndrome). Antibodies and other binding proteins, are used as immunoassay
CC reagents to detect zsig46 or cells expressing it, e.g. for assessing
CC thyroid function to produce anti-idiotypic antibodies, for affinity
CC purification of zsig46, to screen expression libraries, for neutralise
CC zsig46 activity, and to deliver toxins, radioisotopes etc. for
CC therapeutic or diagnostic purposes. Agonists of the product can be used

C to promote growth, differentiation and proliferation of specific cell
C types, e.g. for treating (extra)thyroid diseases or as additive to cell
C cultures
X Sequence 346 AA;

Query Match 95.3%; Score 162; DB 2; Length 346;
Best Local Similarity 96.7%; Pred. No. 7.8e-14;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 MRGAGARGASCMALMLAVPGMS 30
b 1 MRGAGARGASCMALMLAVPGMS 30

RESULT 4
AAU29418
D AAU29418 standard; protein, 192 AA.

AAU29418;

18-DEC-2001 (first entry)

Human G protein-coupled receptor (GPCR) polypeptide #39.

Human; G protein-coupled receptor; GPCR; mental disorder; schizophrenia;
neurological disorder; metabolic disorder; cancer; rheumatoid arthritis;
thyroid disorder; neurodegenerative disorder; cardiovascular disorder;
renal failure; autoimmune disorder; hyperproliferative disorder; HIV;
human immunodeficiency virus; viral infection; neuroprotective;
immunostimulant; neuroleptic; nootropic; tranquiliser; antidepressant;
anorectic; gene therapy.

Homo sapiens.

WO200168858-A2.

20-SEP-2001.

16-MAR-2001; 2001WO-US008456.

16-MAR-2000; 2000US-0187783P.

16-MAR-2000; 2000US-0189907P.

16-MAR-2000; 2000US-0189917P.

16-MAR-2000; 2000US-0189918P.

16-MAR-2000; 2000US-0189960P.

29-MAR-2000; 2000US-0192155P.

29-MAR-2000; 2000US-0192234P.

29-MAR-2000; 2000US-0192830P.

29-MAR-2000; 2000US-0192916P.

29-MAR-2000; 2000US-0192923P.

29-MAR-2000; 2000US-0192933P.

29-MAR-2000; 2000US-0192945P.

(PHAA) PHARMACIA & UPJOHN CO.

Vogeli G;

WPI: 2001-607458/69.

N-PSDB; AAS46857.

Nucleic acid encoding G-protein coupled receptors, useful for the
prevention, diagnosis and treatment of mental disorders.

Claim 31; Page 73; 274pp; English.

Sequences AAU29380-AAU29509 represent human G protein-coupled receptor
(GPCR) polypeptides of the invention. The proteins and the DNA sequences
encoding them can be used to identify compounds which bind to GPCR
polypeptides and in screening for compounds that modulate GPCR activity.
By screening a human subject for the presence of mutations in GPCR DNA,
GPCR-related disorder or a genetic predisposition can be diagnosed. The
sequences can also be used for treatment and prevention of mental

CC disorders such as schizophrenia, neurological disorders such as manic
CC depression, metabolic disorders such as obesity, cancer, rheumatoid
CC arthritis, thyroid disorders such as myxoedema, neurodegenerative
CC disorders such as Parkinson's disease, cardiovascular disorders such as
CC atherosclerosis, renal failure, autoimmune disorders, hyperproliferative
CC disorders such as psoriasis and viral infections such as those caused by
CC HIV

XX Sequence 192 AA;

Query Match 35.8%; Score 60.5; DB 4; Length 192;
Best Local Similarity 39.3%; Pred. No. 2.7;
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

QY 4 GAGARGASCM-----ALALLW 22
Db 35 GRAAVARGPCWCPGCPALLVSIALLW 62

RESULT 5
ABG60706
ID ABG60706 standard; protein, 192 AA.

ABG60706;

13-AUG-2002 (first entry)

Novel G protein coupled receptor (ngPCR-x) #39.

G protein coupled receptor; ngPCR-x; immune response; thyroid disorder;
mental disorder; thyrotoxicosis; myxoedema; inflammatory condition;
Crown's disease; cell differentiation; homeostasis; rheumatoid arthritis;
renal failure; autoimmune disorder; movement disorder; CNS disorder;
viral infection; human immunodeficiency virus; HIV; metabolic disorder;
cardiovascular disorder; diabetes; obesity; anorexia; cardiomyopathy;
proliferative disease; cancer; psoriasis; lung cancer; hormonal disorder;
sexual dysfunction.

Homo sapiens.

US2002058306-A1.

16-MAY-2002.

16-MAR-2001; 2001US-00811284.

16-MAR-2000; 2000US-0189783P.

16-MAR-2000; 2000US-0189907P.

16-MAR-2000; 2000US-0189917P.

16-MAR-2000; 2000US-0189918P.

16-MAR-2000; 2000US-0189960P.

24-MAR-2000; 2000US-0192155P.

27-MAR-2000; 2000US-0192234P.

29-MAR-2000; 2000US-0192830P.

29-MAR-2000; 2000US-0192916P.

29-MAR-2000; 2000US-0192923P.

29-MAR-2000; 2000US-0192933P.

29-MAR-2000; 2000US-0192945P.

(VOGE/) VOGELI G.

Vogeli G;

WPI: 2002-434856/46.

N-PSDB; ABK81635.

New isolated nucleic acid encoding a G protein coupled receptor for
producing the receptor which can induce an immune response in a mammal.
Claim 27; Page 44; 216pp; English.
The invention describes an isolated nucleic acid (I) comprising a

Sequence encoding a portion of a G protein coupled receptor (ngPCR-x).
(1) is used to produce a recombinant ngPCR-x polypeptide. A polypeptide encoded by (1) is used to induce an immune response in a mammal. ngPCR-x is used to identify a compound that binds to it and/or modulates its activity. (1) is used to identify animal homologues of ngPCR-x. (1) can be used to diagnose a human subject as having a brain or genetic predisposition disorder, such as a mental disorder. (1) is used to screen for an ngPCR-x related disorder including thyroid disorders (e.g. thyrotoxicosis, myxoedema), renal failure, inflammatory conditions (e.g. Crohn's disease), diseases related to cell differentiation and homeostasis, rheumatoid arthritis, autoimmune disorders, movement disorders, CNS disorders, viral infections (e.g. Human immunodeficiency virus), metabolic and cardiovascular disorders (e.g. diabetes, obesity, anorexia, cardiomyopathies), proliferative diseases and cancers (e.g. psoriasis, lung cancer), hormonal disorders, sexual dysfunction and hereditary mental disorders in a human patient. A host cell comprising (1) is used to screen for a modulator of ngPCR-x activity. ngPCR-x is used to identify compounds that can treat mental disorders. The polypeptide encoded by (1) is used to purify a G protein from a sample. This is the amino acid sequence of a novel G protein coupled receptor (ngPCR-x) protein described in the invention

Sequence 192 AA;

Query Match 35.8%; Score 60.5; DB 5; Length 192;
Best Local Similarity 39.3%; Pred. No. 2.7;
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

4 GAGAGRGASWCW-----ALALM 22
35 GKAVRGPRPCWPCQPALVSIITLW 62

RESULT 6
AAU14162
D AAU14162 standard; protein; 127 AA.

AAU14162;
24-OCT-2001 (first entry)

Human novel protein #33.
Human; novel protein; Antianaemic; osteopathic; antiinflammatory; immunomodulatory; cytotoxic; neuroprotective; vulnereary; nootropic; anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral; antibacterial; antiallergic; dermatological; haemostatic; antisthmatic; thrombolytic; immunogen; antibody; gene therapy; neurological disorder; Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis; tissue regeneration; immune disorder.

Homo sapiens.

WO200155437-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US002623.

25-JAN-2000; 2000US-00491404.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-451939/48.

N-PSDB; AAS22467.

Isolated polypeptides useful for treating anti-inflammatory diseases, nervous system disorders, and for regenerating bone and cartilage.
Example 4; Page 543-544; 894pp; English.

The invention relates to polynucleotides encoding novel human proteins or their active domains. The polypeptides, polynucleotides and antibodies raised against the polypeptides are used in a method of treatment of a mammal and prevention of disorders caused by the aberrant protein expression or activity. The polypeptides can be used as molecular weight markers, food supplements, and in antibody production. The polypeptides are used to identify compounds which bind to the polypeptides. CC Polynucleotides of the invention are used as probes and primers, for CC sequencing, for chromosome or gene mapping, in the production of CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene CC therapy. Polypeptides of the invention can be used to target drugs to a CC tumour, in assays to determine biological activity, to raise CC antibodies/ elicit an immune response, to determine quantitative protein CC levels, as tissue markers, and to isolate receptors or ligands. CC Polypeptides of the invention may also be useful in treating platelet CC disorders, stem cell disorders, regenerating bone, cartilage, tendon, CC ligament and/or nerve tissue, wound healing, treating burns, promoting CC the proliferation, differentiation and survival of stem cells, as a CC contraceptive, treating osteoporosis and osteoarthritis, anaemia, CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft- CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory CC diseases, nervous system disorders, and infection. The present sequence CC represents a protein of the invention

Sequence 127 AA;

Query Match 35.2%; Score 59.5; DB 4; Length 127;
Best Local Similarity 40.0%; Pred. No. 2.4;
Matches 14; Conservative 3; Mismatches 9; Indels 9; Gaps 2;

3 RGAGARGR-----ASWCALALL--MLAVPRG 28
25 KRGERTGRVTMTKLQWMLGLSTWVALTTG 59

RESULT 7
ABG29252
ID ABG29252 standard; protein; 73 AA.

ABG29252;
18-FEB-2002 (first entry)

Novel human diagnostic protein #29243.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US008631.

31-MAR-2000; 2000US-00540217.

23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

N-PSDB; AAS93439.

New isolated polynucleotide and encoded polypeptides, useful in PT diagnostics, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity.

R 03-DEC-2001; 2001US-0338092P.
R 04-DEC-2001; 2001US-0337185P.
R 03-JAN-2002; 2002US-0345705P.
R 07-MAR-2002; 2002US-00092900.
X
X (CURA-) CTRAGEN CORP.
X
X Padigaru M, Spytek KA, Shenoy SG, Taupier RJ, Pena CEA, Li L,
I Zernhusen BD, Gusev V, Ji W, Gorman L, Miller CE, Kekuda R,
I Patturajan M, Gangoli E, Vernet CAM, Guo X, Tcheuerey V,
I Fernandes ER, Casman SJ, Malayankar UM, Gerlach V, Liu Y, Anderson D,
I Spaderna SK, Catterton E, Burgess C, Lette M, Zhong H, Alsobrook UB,
I Lepley DM, Rieger DK;
X WPI; 2002-723332/78.
R N-PSDB; ABX97157.
R
X NOVX polypeptides and polynucleotides, useful for preventing or treating
T a disorder associated with aberrant NOVX expression or activity e.g.,
T cancer, hypertension, atherosclerosis, cardiomyopathy or bronchial
T asthma.
X
X Claim 1; Page 410; 1103pp; English.
X
X This invention describes novel human NOVX polypeptides which have
C cytosstatic, cardiant, antiarteriosclerotic, antiasthmatic and hypotensive
C activity. Pharmaceutical compositions comprising the NOVX proteins or
C nucleic acid molecules or NOVX antibodies are useful for preventing or
C treating a disorder associated with aberrant NOVX expression or activity
C e.g. cancer, hypertension, atherosclerosis, cardiomyopathy or bronchial
C asthma. The products of the invention can be used for gene therapy or in
C a vaccine. ABU65041-ABU65218 represent the NOVX polypeptides encoded by
C ABX97008-ABX97185
X
X Sequence 403 AA;
Q
Query Match 33.1%; Score 56; DB 5; Length 403;
Best Local Similarity 37.8%; Pred. No. 24;
Matches 14; Conservative 1; Mismatches 14; Indels 8; Gaps 1;
Y
2 RRGAGAGRGASWCALALIMLAVVPG-----WS 30
b 118 RRGASLRGAAGLAWRLFLILAVTLPSIACILYYWS 154
RESULT 12
D AAR41001 standard; protein; 1093 AA.
X
C AAR41001;
X
X 25-MAR-2003 (revised)
T 25-FEB-1994 (first entry)
X
X Human myotonic dystrophy gene protein.
X
X Abnormality: muscular dystrophy; CHR 19; chromosome 19; protein kinase;
W polymerase chain reaction; brain.
X
X Homo sapiens.
X
X
X Key Location/Qualifiers
T Region 1..1093
T /note="encoded by predicted reading frame a, x's in the
T sequence indicate stop codons in the reading frame"
X
X WO9317104-A1.
X
X 02-SEP-1993.
X
X 19-FEB-1993; 93MO-US001545.
X
X 20-FEB-1992; 92US-00839255.
R

XX
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Brook JD, Housman DE;
PI
XX
XX WPI; 1993-288410/36.
DR
XX
XX
PT DNA sequence of myotonic dystrophy gene - used to produce probes and
PT identify CHR 19 abnormality and protein kinase responsible.
XX
XX
PS Disclosure; Fig 6; 64pp; English.
XX
XX The sequence is that encoded by predicted reading frame a of the human
CC myotonic dystrophy (DM) gene. It may be used in the identification of
CC individuals affected by DM. (updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 1093 AA;
QY
Query Match 33.1%; Score 56; DB 2; Length 1093;
Best Local Similarity 39.5%; Pred. No. 67;
Matches 15; Conservative 1; Mismatches 6; Indels 16; Gaps 3;
Db 207 RGHGAACPTGCPRCGXGSSWCWTRA-SW-----GMS 238
3 RGAAGAA-----RGRASWCALALIMLAVVPGWS 30
RESULT 13
AAG93279
ID AAG93279 standard; protein; 371 AA.
XX
XX AAG93279;
AC
XX
DT 13-SEP-2001 (first entry)
XX
XX Human protein HP03145.
DE
XX
XX Human; gene therapy; tumour.
KW
XX
OS Homo sapiens.
XX
XX WO200142302-A1.
PN
XX
PD 14-JUN-2001.
XX
PF 06-DEC-2000; 2000WO-JP008631.
XX
XX 06-DEC-1999; 99JP-00346863.
PR 06-DEC-1999; 99JP-00346864.
PR 08-FEB-2000; 2000JP-00031062.
PR 10-FEB-2000; 2000JP-00034090.
PR 10-FEB-2000; 2000JP-00034091.
PR 14-FEB-2000; 2000JP-00035829.
PR 14-FEB-2000; 2000JP-00035829.
PR 14-MAR-2000; 2000JP-00071161.
PR 30-MAY-2000; 2000JP-00160851.
XX
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA
XX
XX Kato S, Eguchi C, Saeki M;
PI
XX
XX
DR WPI; 2001-381646/40.
DR N-PSDB; AAH68564.
PT Human protein originated from tumor cell line, applicable as drug,
PT reagent for studying intracellular protein networks and protein source
PT for drug screening, also encoded cDNA for gene diagnosis and gene
PT therapy.
XX
XX
PS Claim 1; Page 246-248; 471pp; Japanese.
XX
XX The present sequence is a human protein. The human protein, preferably
CC originated from tumour cell line, is applicable as a drug, a reagent for
CC

infections.

Claim 1; Page 149-150; 170pp; English.

The invention relates to a new isolated human enzyme (NZMS) (I) and the encoding nucleic acid (II). The polypeptides and polynucleotides are useful in assessing the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of NZMS. The NZMS or its fragments are useful in screening compounds for effectiveness as agonist or antagonist of the polypeptides, or in altering the expression of the target polynucleotide and compounds that specifically bind to or modulate the activity of the polypeptide. The antibody is useful in diagnosing a condition or disease associated with the decreased expression or overexpression of NZMS and in detecting (I). (II) is useful in the production of antibodies. The microarray is useful in monitoring or measuring gene expression profiles. The polypeptides and polynucleotides are useful in diagnosing, treating and preventing diseases or conditions associated with the decreased expression or overexpression of NZMS, such as immune system disorders (e.g. acquired immune deficiency syndrome (AIDS), contact dermatitis), immune deficiencies (e.g. common variable immunodeficiency, Chediak-Higashi syndrome), developmental disorders (e.g. Cushing's syndrome, hypothyroidism), eye disorders (e.g. glaucoma, cataract), metabolic disorders (e.g. cystic fibrosis, sickle-cell anaemia), smooth muscle disorders (e.g. asthma, lactic acidosis), neurological disorders (e.g. Parkinson's disease, epilepsy), cardiac disorders (e.g. congestive heart failure, myocardial infarction), pulmonary disorders (emphysema, bronchitis), parasitic infections (amoebiasis, filariasis), and cell proliferative disorders (e.g. cancer, leukaemia). The microarray is useful in monitoring or measuring protein-protein interactions, drug-target interactions, and gene expression profiles. The present sequence represents the amino acid sequence of one of 11 novel human enzymes (NZMS1-11) of the invention

Sequence 371 AA;

Query Match 32.8%; Score 55.5; DB 5; Length 371;
Best Local Similarity 55.6%; Pred. No. 26;
Matches 15; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

3 RGGAARGASWCMALIMLAIVPGW 29
5 RAAGFARGLR---ALALAWL---PGW 24

Search completed: June 8, 2004, 14:03:57
Job time : 8.95833 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

Run on: June 8, 2004, 14:00:33 ; Search time 2.3333 Seconds
(without alignments)
663.762 Million cell updates/sec

Title: US-10-010-050A-2_COPY_1_30
Effect score: 169
Sequence: 1 MRRGAGARGASWCWALMLLWLVPGWS 30

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents.Aa: *
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep: *
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep: *
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep: *
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep: *
5: /cgn2_6/prodata/2/iaa/6C_COMB.pep: *
6: /cgn2_6/prodata/2/iaa/backfiller1.pep: *

Pred. No. is the number of results predicted by chance a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	32.5	109	4	US-09-252-991A-31463 Sequence 31463, A
2	54.5	32.2	263	4	US-09-252-991A-16906 Sequence 16906, A
3	53	31.4	74	4	US-09-621-976-6090 Sequence 6090, App
4	53	31.4	215	3	US-09-220-528-104 Sequence 104, App
5	53	31.4	228	4	US-09-252-991A-32898 Sequence 32898, A
6	52.5	31.1	575	1	US-07-683-957B-2 Sequence 2, Appl
7	52	30.8	170	4	US-09-252-991A-17705 Sequence 16705, A
8	52	30.8	217	4	US-09-252-991A-19862 Sequence 19862, A
9	51.5	30.5	102	4	US-09-252-991A-23296 Sequence 23296, A
10	51.5	30.5	194	4	US-09-252-991A-21199 Sequence 21199, A
11	51.5	30.5	370	4	US-09-252-991A-27810 Sequence 27810, A
12	51	30.2	85	4	US-09-431-705-22 Sequence 22, Appl
13	51	30.2	88	4	US-09-205-258-274 Sequence 274, App
14	51	30.2	1310	4	US-09-170-496D-290 Sequence 290, App
15	51	30.2	1310	4	US-09-364-425B-55 Sequence 55, Appl
16	50.5	29.9	191	2	US-08-469-412A-4 Sequence 4, Appl
17	50.5	29.9	191	2	US-09-021-715-4 Sequence 4, Appl
18	50.5	29.9	209	4	US-08-311-731A-18 Sequence 18, Appl
19	50.5	29.9	301	4	US-09-252-991A-18062 Sequence 18062, A
20	50	29.6	127	4	US-09-252-991A-26303 Sequence 26303, A
21	50	29.6	286	4	US-08-469-260A-275 Sequence 275, App
22	49.5	29.3	88	4	US-08-469-260A-275 Sequence 275, App
23	49.5	29.3	88	4	US-08-469-260A-275 Sequence 275, App
24	49.5	29.3	88	4	US-08-469-260A-275 Sequence 275, App
25	49.5	29.3	92	4	US-09-247-155-98 Sequence 98, Appl
26	49.5	29.3	147	4	US-09-252-991A-23997 Sequence 23997, A
27	49.5	29.3	196	4	US-09-252-991A-31079 Sequence 31079, A

28	49.5	29.3	248	4	US-09-134-001C-5085 Sequence 5085, App
29	49.5	29.3	303	4	US-09-252-991A-17179 Sequence 17179, A
30	49	29.0	718	4	US-09-252-991A-25696 Sequence 25696, A
31	49	29.0	1498	4	US-09-252-991A-31234 Sequence 31234, A
32	48.5	28.7	251	4	US-09-252-991A-27585 Sequence 27585, A
33	48.5	28.7	534	4	US-09-252-991A-17265 Sequence 17265, A
34	48.5	28.4	35	4	US-09-148-545-204 Sequence 204, App
35	48	28.4	170	4	US-09-252-991A-26306 Sequence 26306, A
36	48	28.4	222	4	US-09-148-545-135 Sequence 135, App
37	48	28.4	229	4	US-09-252-991A-22195 Sequence 22195, A
38	48	28.4	260	4	US-09-252-991A-25102 Sequence 25102, A
39	48	28.4	296	4	US-09-489-039A-14080 Sequence 14080, A
40	48	28.4	337	4	US-09-252-991A-19602 Sequence 19602, A
41	48	28.4	439	4	US-09-489-039A-13336 Sequence 13336, A
42	48	28.4	495	4	US-09-489-039A-10528 Sequence 10528, A
43	48	28.4	718	4	US-09-657-960-3 Sequence 3, Appl
44	48	28.4	1421	4	US-09-252-991A-17805 Sequence 17805, A
45	47.5	28.1	191	4	US-09-252-991A-25365 Sequence 25365, A

ALIGNMENTS

```
RESULT 1
US-09-252-991A-31463
; Sequence 31463, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31463
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31463

Query Match      32.5%; Score 55; DB 4; Length 109;
Best Local Similarity 42.1%; Pred. No. 2;
Matches 8; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY      12 ASMCWALMLWLVPGWS 30
Db      27 AAMCWSARSTWACSSAMS 45

RESULT 2
US-09-252-991A-16906
; Sequence 16906, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16906
; LENGTH: 263
; TYPE: PRT
```

ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16906

Query Match 32.2%; Score 54.5; DB 4; Length 263;
Best Local Similarity 46.4%; Pred. No. 5.9;
Matches 13; Conservative 1; Mismatches 7; Indels 7; Gaps 2;

QY 3 RGAGAGRAGSMCWALALMLAVPGWS 30
DB 23 RRGVARGCRARCMR---RW----PGWS 43

RESULT 3
US-09-621-976-6090
Sequence 6090, Application US/09621976
Patent No. 6639063

GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.

APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2

CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335

SOFTWARE: Patent.pm
SEQ ID NO 6090
LENGTH: 74

TYPE: PRT
ORGANISM: Homo sapiens
US-09-621-976-6090

Query Match 31.4%; Score 53; DB 4; Length 74;
Best Local Similarity 39.3%; Pred. No. 2.5;
Matches 11; Conservative 3; Mismatches 10; Indels 4; Gaps 1;

QY 4 GAGAGRGA---SWCWALALMLAVVP 27
DB 13 GLGAAPGRGEPGQGWCMARSIFRCYP 40

RESULT 4
US-09-220-528-104

Sequence 104, Application US/09220528A
Patent No. 6284540

GENERAL INFORMATION:

APPLICANT: Milbrandt, Jeffrey D.
APPLICANT: Balon, Robert H.

TITLE OF INVENTION: Artemin, A No. 6284540e1 Neurotrophic Factor
FILE REFERENCE: 6029-7998

CURRENT APPLICATION NUMBER: US/09/220,528A
CURRENT FILING DATE: 1998-12-24

EARLIER APPLICATION NUMBER: 09/218,698
EARLIER FILING DATE: 1998-12-22

EARLIER APPLICATION NUMBER: 60/108,148
EARLIER FILING DATE: 1998-11-12

EARLIER APPLICATION NUMBER: 09/163,283
EARLIER FILING DATE: 1998-09-29

NUMBER OF SEQ ID NOS: 120
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 104
LENGTH: 215

TYPE: PRT
ORGANISM: Homo sapiens
US-09-220-528-104

Query Match 31.4%; Score 53; DB 3; Length 215;
Best Local Similarity 35.7%; Pred. No. 7.6;
Matches 10; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY 2 RRGAGAGRAGSMCWALALMLAVPGW 29
DB 73 QRRGAGAACRSMWCRGCRARSATPTSW 100

RESULT 5
US-09-252-991A-32898
Sequence 32898, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 32898
LENGTH: 228

TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32898

Query Match 31.4%; Score 53; DB 4; Length 228;
Best Local Similarity 36.7%; Pred. No. 8.1;
Matches 11; Conservative 4; Mismatches 9; Indels 6; Gaps 2;

QY 2 RRGAGA-ARGRASMCWALALMLAVPGWS 30
DB 25 QRRGPHTRGCSATCW----WSAATPAMA 49

RESULT 6
US-07-683-957B-2

Sequence 2, Application US/07683957B
Patent No. 5310880

GENERAL INFORMATION:

APPLICANT: Donahoe, Patricia K.
APPLICANT: Ragin, Richard C.

APPLICANT: MacLaughlin, David T.
TITLE OF INVENTION: Purification of M 11exian Inhibiting

TITLE OF INVENTION: Substance
NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSER: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, N.W.

CITY: Washington
STATE: D.C.

COUNTRY: U.S.A.
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/683,957B
FILING DATE: 19910412

CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:

NAME: Goldstein, Jorge A.
REGISTRATION NUMBER: 29,021

REFERENCE/DOCKET NUMBER: 0609.3060000
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 466-0800
TELEFAX: (202) 833-8716

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 575 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

S-07-683-957B-2

Query Match 31.1%; Score 52.5; DB 1; Length 575;
Best Local Similarity 44.0%; Pred. No. 25;
Matches 11; Conservative 4; Mismatches 7; Indels 3; Gaps 1;

Y 4 GAGAGRAGACWALMLAVPG 28
41 GAGALIFQAMWPLSLML---PG 62

RESULT 7

S-09-252-991A-16705
Sequence 16705; Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 16705

LENGTH: 170

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

S-09-252-991A-16705

Query Match 30.8%; Score 52; DB 4; Length 170;
Best Local Similarity 39.3%; Pred. No. 8,2;
Matches 11; Conservative 1; Mismatches 8; Indels 8; Gaps 1;

Y 3 RAGAGARGASW-----CVALALLW 22
108 RAAGATRRRRGWTCCANPRPCWSAGSW 135

RESULT 8
S-09-252-991A-19862
Sequence 19862; Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 19862

LENGTH: 217

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

S-09-252-991A-19862

Query Match 30.8%; Score 52; DB 4; Length 217;
Best Local Similarity 61.1%; Pred. No. 11;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Y 2 RAGAGARGASWCVALLA 19
191 RRGAGRGGRGACGCGA 208

RESULT 9

US-09-252-991A-23296
Sequence 23296; Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 23296

LENGTH: 102

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-23296

Query Match 30.5%; Score 51.5; DB 4; Length 102;
Best Local Similarity 39.5%; Pred. No. 5,6;
Matches 15; Conservative 5; Mismatches 9; Indels 9; Gaps 3;

Y 1 MRGAGARGASW-C--WALMLAVP-----GW 29
Db 48 LRLAIFSAAGRSAYFCSNRWICATWLAKPSMIRLW 85

RESULT 10

US-09-252-991A-21199
Sequence 21199; Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 21199

LENGTH: 194

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-21199

Query Match 30.5%; Score 51.5; DB 4; Length 194;
Best Local Similarity 40.6%; Pred. No. 11;
Matches 13; Conservative 4; Mismatches 12; Indels 3; Gaps 1;

Y 2 RAGAGARGASWCVALLML---AVYPGMS 30
Db 4 RAGAGRCRGGRGCPVALSVHPLRPGELPAWS 35

RESULT 11

US-09-252-991A-27810
Sequence 27810; Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 27810
LENGTH: 370
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27810

Query Match 30.5%; Score 51.5; DB 4; Length 370;
Best Local Similarity 38.5%; Pred. No. 22;
Matches 10; Conservative 5; Mismatches 8; Indels 3; Gaps 1;

2y 7 AARGRAS--WCMALALMLAVPGM 29
Db 112 AANSRPSATWMEFSATMTLTLAV 137

RESULT 12
US-09-431-705-22
Sequence 22, Application US/09431705
Patent No. 6585975
GENERAL INFORMATION:
APPLICANT: Kleantous, Harold
APPLICANT: Londono-Arcila, Patricia
APPLICANT: Freeman, Donna
TITLE OF INVENTION: Use of salmonella vectors for
FILE REFERENCE: 06132/060001
CURRENT APPLICATION NUMBER: US/09/431,705
CURRENT FILING DATE: 1999-11-01
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 22
LENGTH: 85
TYPE: PRT
ORGANISM: Escherichia coli
US-09-431-705-22

Query Match 30.2%; Score 51; DB 4; Length 85;
Best Local Similarity 52.9%; Pred. No. 5.4;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

2y 11 RASWCMALALMLAVVP 27
Db 16 QASWCHARRLVWLSAP 32

RESULT 13
US-09-205-258-274
Sequence 274, Application US/09205258
Patent No. 6525174
GENERAL INFORMATION:
APPLICANT: Young et al.
TITLE OF INVENTION: 207 Human Secreted Proteins
FILE REFERENCE: P2007P1
CURRENT APPLICATION NUMBER: US/09/205,258
CURRENT FILING DATE: 1998-12-04
EARLIER APPLICATION NUMBER: PCT/US98/11422
EARLIER FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048,885
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,375
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,881
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,880
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,896
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,020

EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,896
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,895
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,884
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,894
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,971
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,882
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,899
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,893
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,900
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,901
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,915
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,019
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,972
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,916
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,373
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,875
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,374
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,917
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,949
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,883
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,897
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,898
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,962
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,963
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,877
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,878
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/070,923
EARLIER FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/092,921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/094,657
EARLIER FILING DATE: 1998-07-30
NUMBER OF SEQ ID NOS: 1227
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 274
LENGTH: 88
TYPE: PRT
ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: SITE
LOCATION: (53)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (88)
OTHER INFORMATION: Xaa equals stop translation
S-09-205-258-274

Query Match 30.2%; Score 51; DB 4; Length 88;
Best Local Similarity 56.5%; Pred. No. 5.6;
Matches 13; Conservative 1; Mismatches 7; Indels 2; Gaps 1;

Y 1 MRRGA--GAAGRASMCALAL 21
b 44 MAKGRGAGKRRRNGIAYTL 66

ESULT 14
S-09-170-496D-290
Sequence 290, Application US/09170496D
Patent No. 6555339

GENERAL INFORMATION:

APPLICANT: Behan, Dominic P.

APPLICANT: Chalmers, Derek T.

APPLICANT: Liaw, Chen W.

TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-

FILE REFERENCE: AREN-0040

CURRENT APPLICATION NUMBER: US/09/170,496D

CURRENT FILING DATE: 1998-10-13

NUMBER OF SEQ ID NOS: 294

SOFTWARE: PatentIn version 3.1

SEQ ID NO 230

LENGTH: 1310

TYPE: PRT

ORGANISM: Homo sapiens

S-09-170-496D-290

Query Match 30.2%; Score 51; DB 4; Length 1310;
Best Local Similarity 52.9%; Pred. No. 95;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Y 11 RASMCWALALMLAVVP 27
b 907 QASWCHARRLWLHSAP 923

ESULT 15
S-09-364-425B-55

Sequence 55, Application US/09364425B

Patent No. 6653086

GENERAL INFORMATION:

APPLICANT: Behan, Dominic P.

APPLICANT: Chalmers, Derek T.

APPLICANT: Liaw, Chen W.

APPLICANT: Lin, I-Lin

APPLICANT: Lowitz, Kevin P.

APPLICANT: Chen, Ruoping

TITLE OF INVENTION: Endogenous, Constitutively Activated G Protein-Coupled Orphan Rece

FILE REFERENCE: AREN0047

CURRENT APPLICATION NUMBER: US/09/364,425B

CURRENT FILING DATE: 2001-12-18

PRIOR APPLICATION NUMBER: 60/094,879

PRIOR FILING DATE: 1998-07-31

PRIOR APPLICATION NUMBER: 60/106,300

PRIOR FILING DATE: 1998-10-30

PRIOR APPLICATION NUMBER: 60/110,906

PRIOR FILING DATE: 1998-12-04

PRIOR APPLICATION NUMBER: 60/121,851

PRIOR FILING DATE: 1999-02-26

NUMBER OF SEQ ID NOS: 60

SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55
; LENGTH: 1310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-364-425B-55

Query Match 30.2%; Score 51; DB 4; Length 1310;
Best Local Similarity 52.9%; Pred. No. 95;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 11 RASMCWALALMLAVVP 27
Db 907 QASWCHARRLWLHSAP 923

Search completed: June 8, 2004, 14:09:16
Job time: 2.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

Run on: June 8, 2004, 14:04:04 ; Search time 6.16667 Seconds
(without alignments)
1368.675 Million cell updates/sec

File: US-10-010-050A-2_COPY_1_30
Effect score: 169
Sequence: 1 MRRGAGARGASGCMALALMLAVVPGMS 30

Coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Seached: 1155919 seqs, 28138677 residues

total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	169	100.0	346	9	US-09-122-383-2
2	169	100.0	346	12	US-10-653-595-126
3	169	100.0	346	13	US-10-010-050A-2
4	169	100.0	346	12	US-09-397-945-126
5	62	36.7	588	14	US-10-156-761-11538
6	60.5	35.8	192	9	US-09-811-284-167
7	59.5	35.2	127	15	US-10-291-265-269
8	58	34.3	188	9	US-09-989-920-240
9	57	33.7	19608	15	US-10-084-846A-8
10	56	33.1	214	14	US-10-080-170-198
11	56	33.1	403	12	US-10-092-900A-300
12	55.5	32.8	371	10	US-09-890-688-60
13	53	31.4	28	9	US-09-242-450A-13
14	53	31.4	65	9	US-09-242-450A-3
15	53	31.4	117	15	US-10-094-749-2218

16	53	31.4	160	14	US-10-156-761-10933	Sequence 10933, A
17	53	31.4	215	9	US-09-220-920-104	Sequence 104, Appl
18	53	31.4	346	12	US-10-047-825-10	Sequence 10, Appl
19	53	31.4	368	12	US-10-425-114-55731	Sequence 55731, A
20	52.5	31.1	199	12	US-10-221-278-727	Sequence 727, Appl
21	52.5	31.1	199	15	US-10-291-172-727	Sequence 727, Appl
22	52.5	31.1	333	9	US-09-738-626-4771	Sequence 4771, Ap
23	52.5	31.1	360	12	US-10-282-122A-50396	Sequence 50396, A
24	52.5	31.1	484	12	US-10-424-599-207280	Sequence 207280, A
25	52	30.8	82	12	US-10-424-599-167251	Sequence 167251, A
26	52	30.8	560	10	US-09-863-776-69	Sequence 69, Appl
27	52	30.8	599	15	US-10-104-047-2306	Sequence 2306, Ap
28	52	30.8	653	10	US-09-863-776-28	Sequence 28, Appl
29	52	30.8	658	9	US-09-759-056-5	Sequence 5, Appl
30	52	30.8	658	9	US-09-901-812-5	Sequence 5, Appl
31	52	30.8	658	10	US-09-863-776-68	Sequence 68, Appl
32	52	30.8	662	10	US-09-863-776-30	Sequence 30, Appl
33	52	30.8	667	9	US-09-759-056-2	Sequence 2, Appl
34	52	30.8	667	9	US-09-901-812-2	Sequence 2, Appl
35	52	30.8	667	10	US-09-863-776-32	Sequence 32, Appl
36	52	30.8	667	10	US-09-863-776-67	Sequence 67, Appl
37	52	30.8	667	12	US-10-219-535-80	Sequence 80, Appl
38	52	30.8	667	12	US-10-232-230-80	Sequence 80, Appl
39	52	30.8	667	12	US-10-232-224-80	Sequence 80, Appl
40	52	30.8	667	14	US-10-227-884-80	Sequence 80, Appl
41	52	30.8	667	14	US-10-230-163-80	Sequence 80, Appl
42	52	30.8	667	14	US-10-230-338-80	Sequence 80, Appl
43	52	30.8	667	14	US-10-218-631-80	Sequence 80, Appl
44	52	30.8	667	14	US-10-230-414-80	Sequence 80, Appl
45	52	30.8	667	14	US-10-216-159A-80	Sequence 80, Appl

ALIGNMENTS

```
RESULT 1
US-09-122-383-2
; Sequence 2, Application US/09122383A
; Patent No. US20020042093A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
; FILE REFERENCE: 97-38
; CURRENT APPLICATION NUMBER: US/09/122,383A
; CURRENT FILING DATE: 1998-07-24
; EARLIER APPLICATION NUMBER: 60/053,613
; EARLIER FILING DATE: 1997-07-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-122-383-2

Query Match      100.0%; Score 169; DB 9; Length 346;
Best Local Similarity 100.0%; Pred. No. 3.8e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MRRGAGARGASGCMALALMLAVVPGMS 30
Db      1 MRRGAGARGASGCMALALMLAVVPGMS 30

RESULT 2
US-10-653-595-126
; Sequence 126, Application US/10653595
; Publication No. US20040048304A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 95 Human secreted proteins
```

```
FILE REFERENCE: P2027P1
CURRENT APPLICATION NUMBER: US/10/653,595
CURRENT FILING DATE: 2003-09-03
PRIOR APPLICATION NUMBER: US 09/397945
PRIOR FILING DATE: 1999-09-17
PRIOR APPLICATION NUMBER: PCT/US99/05804
PRIOR FILING DATE: 1998-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 126
LENGTH: 346
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (242)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (246)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
S-10-653-595-126
Query Match          100.0%; Score 169; DB 12; Length 346;
Best Local Similarity 100.0%; Pred. No. 3,8e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ly      1 MRRGAGARGRASWCWALALMLAVVPGWS 30
      1 MRRGAGARGRASWCWALALMLAVVPGWS 30

b      1 MRRGAGARGRASWCWALALMLAVVPGWS 30
      1 MRRGAGARGRASWCWALALMLAVVPGWS 30

RESULT 3
S-10-010-050A-2
Sequence 2, Application US/10010050A
Publication No. US20020173624A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Gilbertson, Debra G.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
FILE REFERENCE: 97-38C1
CURRENT APPLICATION NUMBER: US/10/010,050A
CURRENT FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: US 09/122,383
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: US 60/053,613
PRIOR FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 346
TYPE: PRT
ORGANISM: Homo sapien
S-10-010-050A-2
```

```
Query Match          100.0%; Score 169; DB 13; Length 346;
Best Local Similarity 100.0%; Pred. No. 3,8e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MRRGAGARGRASWCWALALMLAVVPGWS 30
      1 MRRGAGARGRASWCWALALMLAVVPGWS 30

Db      1 MRRGAGARGRASWCWALALMLAVVPGWS 30
      1 MRRGAGARGRASWCWALALMLAVVPGWS 30

RESULT 4
US-09-397-945-126
Sequence 126, Application US/09397945
Publication No. US20030065139A1
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc. et al.
TITLE OF INVENTION: 95 Human secreted proteins
FILE REFERENCE: P2027P1
CURRENT APPLICATION NUMBER: US/09/397,945
CURRENT FILING DATE: 1999-09-17
PRIOR APPLICATION NUMBER: PCT/US99/05804
PRIOR FILING DATE: 1998-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,581
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,577
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,563
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,313
PRIOR FILING DATE: 1998-04-01
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 126
LENGTH: 347
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (242)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (246)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (347)
OTHER INFORMATION: Xaa equals stop translation
US-09-397-945-126
Query Match          100.0%; Score 169; DB 12; Length 347;
Best Local Similarity 100.0%; Pred. No. 3,8e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MRRGAGARGRASWCWALALMLAVVPGWS 30
      1 MRRGAGARGRASWCWALALMLAVVPGWS 30

Db      1 MRRGAGARGRASWCWALALMLAVVPGWS 30
      1 MRRGAGARGRASWCWALALMLAVVPGWS 30

RESULT 5
```

S-10-156-761-11538
Sequence 11538, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 11538
LENGTH: 588
TYPE: PRT
ORGANISM: Streptomyces avermitilis
3-10-156-761-11538
Query Match 36.7%; Score 62; DB 14; Length 588;
Best Local Similarity 28.6%; Pred. No. 23;
Matches 14; Conservative 5; Mismatches 8; Indels 22; Gaps 1;
Y 3 RGAGAR-----GRASWCALALLMLAVVPG 29
167 RGAGSRRTLYATKNTLLIGVYGRYGDKSSWSWAGGLIMLAPR 215
3-09-811-284-167
Sequence 167, Application US/09811284
Patent No. US20020058306A1
GENERAL INFORMATION:
APPLICANT: Vogel, Gabriel
TITLE OF INVENTION: No. US20020058306A1 G Protein-coupled Receptors
FILE REFERENCE: 00167U51
CURRENT APPLICATION NUMBER: US/09/811,284
CURRENT FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/189,783
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,907
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,918
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,960
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,917
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/192,945
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,916
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,923
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,933
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,830
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,234
PRIOR FILING DATE: 2000-03-27
PRIOR APPLICATION NUMBER: 60/192,155
PRIOR FILING DATE: 2000-03-24
PRIOR APPLICATION NUMBER: 60/192,935
PRIOR FILING DATE: 2000-03-29
NUMBER OF SEQ ID NOS: 258
SOFTWARE: PatentIn version 3.0

SEQ ID NO 167
LENGTH: 192
TYPE: PRT
ORGANISM: Homo sapiens
US-09-811-284-167
Query Match 35.8%; Score 60.5; DB 9; Length 192;
Best Local Similarity 39.3%; Pred. No. 13;
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;
Y 4 GAGANGRASWC-----ALALL 22
35 GAAVVRGRCWCPCCPALLVSIALLW 62
RESULT 7
US-10-291-265-269
Sequence 269, Application US/10291265
Publication No. US20030232054A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
APPLICANT: Tang et al
TITLE OF INVENTION: No. US20030232054A1 Nucleic Acids and Polypeptides
FILE REFERENCE: 21272-017 (785)
CURRENT APPLICATION NUMBER: US/10/291,265
CURRENT FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: 09/617,746
PRIOR FILING DATE: 2000-07-17
PRIOR APPLICATION NUMBER: 09/631,451
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 09/633,870
PRIOR FILING DATE: 2000-09-15
NUMBER OF SEQ ID NOS: 944
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 269
LENGTH: 127
TYPE: PRT
ORGANISM: Homo sapiens
US-10-291-265-269
Query Match 35.2%; Score 59.5; DB 15; Length 127;
Best Local Similarity 40.0%; Pred. No. 12;
Matches 14; Conservative 3; Mismatches 9; Indels 9; Gaps 2;
Y 3 RGAGAR-----ASWCALALL--WLA VPG 28
25 RRGRTGRVMTKLAQWLWGIALIGSTVALTTG 59
RESULT 8
US-09-989-920-240
Sequence 240, Application US/09989920
Patent No. US20020172957A1
GENERAL INFORMATION:
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Heve
APPLICANT: Chen, Sei-Yu
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prote
FILE REFERENCE: DEX-0291
CURRENT APPLICATION NUMBER: US/09/989,920
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/252,500
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 284
SOFTWARE: PatentIn version 3.1
SEQ ID NO 240
LENGTH: 188
TYPE: PRT
ORGANISM: Homo sapien

US-09-989-920-240

Query Match 34.3%; Score 58; DB 9; Length 188;
Best Local Similarity 38.7%; Pred. No. 26;
Matches 12; Conservative 1; Mismatches 12; Indels 6; Gaps 1;

QY 4 GAGARGASWCM-----ALALLMLAVPG 28
DB 19 GGGGGAGAGSWVMGSGGAGALWVAVGCG 49

RESULT 9
US-10-084-846A-8
; Sequence 8, Application US/10084846A
; Publication No. US20040006026A1

GENERAL INFORMATION:

APPLICANT: WEINHAUER, GABRIELE

APPLICANT: MUEHLNBERG, AGNES

APPLICANT: TREFFER, AXEL

APPLICANT: BECHTHOLD, ANDREAS

TITLE OF INVENTION: AVILAMYCIN DERIVATIVES

FILE REFERENCE: 1974-005

CURRENT APPLICATION NUMBER: US/10/084,846A

PRIOR FILING DATE: 2003-02-25

PRIOR APPLICATION NUMBER: PCT/EP01/09815

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: DE 101 09 166.4

NUMBER OF SEQ ID NOS: 120

SOFTWARE: Patent In Ver. 3.2

SEQ ID NO 8

LENGTH: 19608

TYPE: PRT

ORGANISM: Streptomyces viridochromogenes

FEATURE:
OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.

US-10-084-846A-8

Query Match 33.7%; Score 57; DB 15; Length 19608;
Best Local Similarity 42.9%; Pred. No. 1,8e+03;
Matches 12; Conservative 3; Mismatches 11; Indels 2; Gaps 1;

Y 2 RRGAGARGASWCMALLMLAVP 27
DB 2876 RRAAVTPGARGRWVCGPVGWLCADP 2903

RESULT 10
US-10-080-170-198
; Sequence 198, Application US/10080170
; Publication No. US20030129601A1

GENERAL INFORMATION:

APPLICANT: COLE, S.T.

TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR

TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR

TREATMENT OF MYCOBACTERIOSES

FILE REFERENCE: 03495.0218

CURRENT APPLICATION NUMBER: US/10/080,170

PRIOR FILING DATE: 2002-06-10

PRIOR APPLICATION NUMBER: 60/270,123

PRIOR FILING DATE: 2001-02-22

NUMBER OF SEQ ID NOS: 652

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 198

LENGTH: 214

TYPE: PRT

ORGANISM: Mycobacterium leprae

US-10-080-170-198

Query Match 33.1%; Score 56; DB 14; Length 214;
Best Local Similarity 41.7%; Pred. No. 50;
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 3;

QY 3 RGAGARGASWCMALLMLAVP 26
DB 175 RTSALARGITQWRMSIALMLFTIV 198

RESULT 11
US-10-092-900A-300
; Sequence 300, Application US/10092900A
; Publication No. US20040043382A1

GENERAL INFORMATION:

APPLICANT: Padigar, Muralidhara

APPLICANT: Spytek, Kimberly A.

APPLICANT: Shenoy, Suresh G.

APPLICANT: Taupier Jr., Raymond J.

APPLICANT: Pena, Carol E.A.

APPLICANT: Li, Li

APPLICANT: Zethusen, Bryan D.

APPLICANT: Gusev, Vladimil Y.

APPLICANT: Ji, Weizhen

APPLICANT: Gorman, Linda

APPLICANT: Miller, Charles E.

APPLICANT: Kekuda, Ramesh

APPLICANT: Paturajan, Weera

APPLICANT: Gangoli, Esna A.

APPLICANT: Vermet, Corine A.M.

APPLICANT: Guo, Xiaojia Sasha

APPLICANT: Tcherenev, Velizar T.

APPLICANT: Fernandes, Elma R.

APPLICANT: Casman, Stacie J.

APPLICANT: Malyankar, Uriel M.

APPLICANT: Gerlach, Valerie

APPLICANT: Liu, Yi

APPLICANT: Anderson, David W.

APPLICANT: Spaderna, Steven K.

APPLICANT: Catterton, Elina

APPLICANT: Leite, Mario W.

APPLICANT: Zhong, Haihong

APPLICANT: Alsobrook, John P.

APPLICANT: Lepley, Denise M.

APPLICANT: Rieger, Daniel K.

APPLICANT: Burgess, Catherine E.

TITLE OF INVENTION: No. US20040043382A1e1 Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-290C

CURRENT APPLICATION NUMBER: US/10/092,900A

CURRENT FILING DATE: 2002-03-07

PRIOR APPLICATION NUMBER: USSN 60/274,322

PRIOR FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: USSN 60/283,675

PRIOR FILING DATE: 2001-04-13

PRIOR APPLICATION NUMBER: USSN 60/338,092

PRIOR FILING DATE: 2001-12-03

PRIOR APPLICATION NUMBER: USSN 60/274,281

PRIOR FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: USSN 60/274,191

PRIOR FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: USSN 60/325,681

PRIOR FILING DATE: 2001-09-27

PRIOR APPLICATION NUMBER: USSN 60/304,354

PRIOR FILING DATE: 2001-07-10

PRIOR APPLICATION NUMBER: USSN 60/279,995

PRIOR FILING DATE: 2001-03-30

PRIOR APPLICATION NUMBER: USSN 60/294,899

PRIOR FILING DATE: 2001-05-31

PRIOR APPLICATION NUMBER: USSN 60/287,424

PRIOR FILING DATE: 2001-04-30

Remaining Prior Application data removed - See File Wrapper or PAM.

NUMBER OF SEQ ID NOS: 768

SEQ ID NO 300

LENGTH: 403

TYPE: PRT

ORGANISM: Homo sapiens

US-10-092-900A-300

Query Match 33.1%; Score 56; DB 12; Length 403;
Best Local Similarity 37.8%; Pred. No. 86;
Matches 14; Conservative 1; Mismatches 14; Indels 8; Gaps 1;

2 RRGAGAGRASWCALALMLAVPG-----WS 30
118 RRGALRGASLGRSLMRLPLFLAVTLPLSLCILITYWS 154

RESULT 12
US-09-890-688-60

Sequence 60, Application US/09890688
Publication No. US20030144475A1
GENERAL INFORMATION:

APPLICANT: Seishi KATO
APPLICANT: Chikashi EGUCHI
TITLE OF INVENTION: Human Proteins and cDNAs thereof
FILE REFERENCE: 2001-1102A/WMC/00653
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: JP 11-346863
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: JP 11-34684
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: JP 2000-31062
PRIOR FILING DATE: 2000-02-08
PRIOR APPLICATION NUMBER: JP 2000-34091
PRIOR FILING DATE: 2000-02-10
PRIOR APPLICATION NUMBER: JP 2000-34090
PRIOR FILING DATE: 2000-02-10
PRIOR APPLICATION NUMBER: JP 2000-35829
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: JP 2000-35899
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: JP 2000-71161
PRIOR FILING DATE: 2000-03-14
PRIOR APPLICATION NUMBER: JP 2000-160851
PRIOR FILING DATE: 2000-05-30
NUMBER OF SEQ ID NOS: 160
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 60

LENGTH: 371
TYPE: PRT

ORGANISM: Homo sapiens

S-09-890-688-60

Query Match 32.8%; Score 55.5; DB 10; Length 371;
Best Local Similarity 55.6%; Pred. No. 91;
Matches 15; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

3 RRGAGAGRASWCALALMLAVPGM 29
5 RAAGFARGLR---ALALML---PQM 24

RESULT 13
S-09-242-450A-13

Sequence 13, Application US/09242450A
Patent No. US2002015552A1
GENERAL INFORMATION:

APPLICANT: Bergmann, Johanna
TITLE OF INVENTION: "PRIONINS", HIGHLY SPECIFIC MARKERS FOR NONINVASIVE PRE-SYMPTOM
TITLE OF INVENTION: DETECTION OF TSE DISEASES AND TARGETS FOR THERAPEUTIC REAGENTS
FILE REFERENCE: 830006-2001
CURRENT FILING DATE: US/09/242,450A
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.0
SEQ ID NO 13

LENGTH: 28

TYPE: PRT
ORGANISM: homo sapiens
FEATURE:
NAME/KEY: TRANSMEM
LOCATION: (1)..(28)
OTHER INFORMATION: membrane spanning helix of CYS from amino acid 25 to amino acid
US-09-242-450A-13

Query Match 31.4%; Score 53; DB 9; Length 28;
Best Local Similarity 45.5%; Pred. No. 20;
Matches 10; Conservative 1; Mismatches 9; Indels 2; Gaps 1;

10 GRASWCW--ALALMLAVPGM 29
4 GAASWWMLGASWWMLGAAPRW 25

RESULT 14

US-09-242-450A-3
Sequence 3, Application US/09242450A
Patent No. US2002015552A1
GENERAL INFORMATION:

APPLICANT: Bergmann, Johanna
TITLE OF INVENTION: "PRIONINS", HIGHLY SPECIFIC MARKERS FOR NONINVASIVE PRE-SYMPTOM
TITLE OF INVENTION: DETECTION OF TSE DISEASES AND TARGETS FOR THERAPEUTIC REAGENTS
FILE REFERENCE: 830006-2001
CURRENT FILING DATE: US/09/242,450A
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 65
TYPE: PRT
ORGANISM: homo sapiens
US-09-242-450A-3

Query Match 31.4%; Score 53; DB 9; Length 65;
Best Local Similarity 45.5%; Pred. No. 41;
Matches 10; Conservative 1; Mismatches 9; Indels 2; Gaps 1;

10 GRASWCW--ALALMLAVPGM 29
28 GAASWWMLGASWWMLGAAPRW 49

RESULT 15

US-10-094-749-2218
Sequence 2218, Application US/10094749
Publication No. US20030219741A1
GENERAL INFORMATION:

APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHKO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749

```

; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2218
; LENGTH: 117
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-094-749-2218

```

```

Query Match
Best Local Similarity 31.4%; Score 53; DB 15; Length 117;
Matches 15; Conservative 2; Mismatches 12; Indels 2; Gaps 2;
2y 2 RRGAGA---ARGRAS-----WCWALALIMLAVPGWS 30
2b 54 RRGAGATGARGRGHGHSSLIAGPWRMLLSFFPKVNNWS 94

```

Search completed: June 8, 2004, 14:11:57
 Job time : 6.16667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: June 8, 2004, 13:59:38 ; Search time 2.16667 Seconds

(without alignments)
1331.883 Million cell updates/sec

title: US-10-010-050A-2_COPY_1_30

effect score: 169
equence: 1 MRRGAGARGRASWCWALALMLAVPGWS 30

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

earched: 283366 seqs, 96191526 residues

otal number of hits satisfying chosen parameters: 283366

imum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	39.1	606	2 C84390	sulfate transport
2	59	34.9	412	2 D95941	conserved hypothet
3	56	33.1	214	2 T44701	probable integral
4	55	32.5	232	2 F82729	conserved hypothet
5	53.5	31.7	159	2 A48428	homeotic protein H
6	53.5	31.7	284	2 A82729	phosphatidate cycli
7	53	31.4	398	2 S11800	pullulanase secret
8	53	31.4	428	2 A12860	conserved hypothet
9	53	31.4	428	2 H97637	hypothetical prote
10	52.5	31.1	220	2 AG3627	mallose transport
11	52.5	31.1	575	2 WFBOM	mullerian inhibiti
12	52	30.8	641	2 A45054	probable intercell
13	51.5	30.5	686	2 S30075	ferric reductase (
14	51.5	30.5	1146	2 A38587	collagen, cornea-s
15	51	30.2	410	2 E95984	conserved hypothet
16	50.5	29.9	210	2 E70671	hypothetical prote
17	50	29.6	106	2 A29760	hypothetical prote
18	50	29.6	163	2 S73039	hypothetical prote
19	50	29.6	201	2 D81944	probable integral
20	50	29.6	219	2 S70311	hypothetical prote
21	50	29.6	221	2 G83964	hypothetical prote
22	50	29.6	249	2 B84147	ABC transporter (p
23	50	29.6	363	2 AD1197	ABC transporter tr
24	50	29.6	372	2 T34631	probable integral
25	50	29.6	417	2 AF3448	hypothetical membr
26	50	29.6	514	2 S17958	cytochrome oxidase
27	50	29.6	514	2 F90770	probable third cyt
28	50	29.6	514	2 B85633	probable third cyt
29	50	29.6	529	2 AF3059	cytochrome d oxida

30	50	29.6	529	2 B98227	cytochrome d oxida
31	49.5	29.3	310	2 T16233	hypothetical prote
32	49.5	29.3	519	2 S75570	apolipoprotein N-a
33	49	29.0	118	2 S52855	hypothetical prote
34	49	29.0	139	2 A71123	hypothetical prote
35	49	29.0	188	2 B82183	ankb protein VC158
36	49	29.0	250	2 AB1555	ABC transporter tr
37	49	29.0	317	2 T35010	probable integral
38	49	29.0	755	2 B75346	probable competenc
39	49	29.0	1221	2 E83327	conserved hypothet
40	48.5	28.7	111	2 T14306	glycine-rich prote
41	48.5	28.7	125	2 C83138	hypothetical prote
42	48.5	28.7	210	2 S36297	T-cell receptor ga
43	48.5	28.7	276	2 B82243	hypothetical prote
44	48.5	28.7	435	2 A12082	hypothetical prote
45	48.5	28.7	688	2 B97152	probable membrane

ALIGNMENTS

RESULT 1

C84390 sulfate transport system permease protein [imported] - Halobacterium sp. NRC-1

C/Species: Halobacterium sp. NRC-1

C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C/Accession: C84390

R/ing, M.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebdhardt, H.; Lowe, T.M.; L 

A/Title: Genome sequence of Halobacterium species NRC-1.

A/Reference number: A81160; MUID:20504483; EUID:11016950

A/Accession: C84390

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-606 <STO>

A/Cross-references: GB:AB004437; NID:g10581803; PID:AA620487.1; GSPDB:GN00138

C/Genetics: cyr11

Query Match

Best Local Similarity 39.1%; Score 66; DB 2; Length 606;

Matches 14; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 3 RGAGARGRASWCWALALMLAVPG 28

DB 508 RALGASRGALWDELPLWPGVAVG 533

RESULT 2

D95941 conserved hypothetical membrane protein, paralogue of Y20848 Smb21292 [imported] - Sino

C/Species: Sinorhizobium meliloti

C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C/Accession: D95941

R/Finan, T.W.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.D.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A/Title: The complete sequence of the 1,663-Kb pSymB megaplasmid from the N2-fixing endo

A/Reference number: A55842; MUID:21396508; PMID:11481431

A/Accession: D95941

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-412 <KOR>

A/Cross-references: GB:AL591985; PID:CA949196.1; PID:g15140681; GSPDB:GN00167

A/Experimental source: strain 1021, megaplasmid pSymB

R/Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Kemp, C.; Lelaure,

hebaul, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C

A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:2136824; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SMD21292
 A:Genome: plasmid

Query Match 34.9%; Score 59; DB 2; Length 412;
 Best Local Similarity 37.0%; Pred. No. 3.1;
 Matches 10; Conservative 6; Mismatches 5; Indels 6; Gaps 1;
 Db 7 7 ARGRASMCWA-----LALLMLAVP 27
 276 ANKGRPTWLTATVLTALFIMLSIVP 302

RESULT 3
 144701
 Probable integral membrane protein (imported) - *Mycobacterium leprae*
 C:Species: *Mycobacterium leprae*
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 28-Jul-2000
 C:Accession: 144701
 R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 Submitted to the EMBL Data Library, May 1998
 A:Reference number: Z22830
 A:Accession: T44701
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-214 <PAR>
 A:Cross-references: EMBL:AL023635; PIDN:CAA19188.1
 A:Experimental source: cosmid B1243
 A:Genetics:
 A>Note: MLCB1243.07
 A:Superfamily: *Mycobacterium tuberculosis* hypothetical protein Ry2968c

Query Match 33.1%; Score 56; DB 2; Length 214;
 Best Local Similarity 41.7%; Pred. No. 4.4;
 Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 Db 3 3 EGAGARGRASCWALALMLAVV 26
 175 RTSALARGIYQWRFSIATLMTIV 138

RESULT 4
 182729
 Conserved hypothetical protein XF1054 [imported] - *Xylella fastidiosa* (strain 955c)
 C:Species: *Xylella fastidiosa*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
 C:Accession: F82729
 R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequence
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: F82729
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-232 <SIM>
 A:Cross-references: GB:AE003942; GB:AE003849; NID:99105990; PIDN:AAF83864.1; GSPDB:GN001
 A:Experimental source: strain 955c
 A:Authors: A.J.G. Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 rtiens, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Caniaro, L.E.A.; Carraro, D.M.; Carre, H
 S-Neto, E.; Docena, C.; El-Dorzy, H.; Facincani, A.P.; Ferreira, A.U.S.
 Submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
 D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre
 hado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 ; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF1054
 C:Superfamily: *Streptomyces coelicolor* probable integral membrane protein SC6G10.12

Query Match 32.5%; Score 55; DB 2; Length 232;
 Best Local Similarity 44.0%; Pred. No. 6.3;
 Matches 11; Conservative 5; Mismatches 7; Indels 2; Gaps 1;
 Db 8 8 ARGRASMCWALALMLAVPG--WS 30
 13 SRNRPWTMLAVPLIFAVLIFLWS 37

RESULT 5
 A48428
 Homeotic protein Hox 1.7 (clone MH-1) - guinea pig (fragment)
 C:Species: *Cavia porcellus* (guinea pig)
 C:Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 17-Oct-1997
 C:Accession: A48428
 R:Rubin, M.R.; Nguyen-Huu, M.C.
 DNA Seq. 1, 115-124, 1990
 A:Title: Alternatively spliced Hox-1.7 transcripts encode different protein products.
 A:Reference number: A48428; MUID:92190539; PMID:1983703
 A:Accession: A48428
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-159 <RUB>
 A>Note: sequence extracted from NCBI backbone (NCBI:89505, NCBI:89507)
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:94-150/Domain: homeobox homology <HDX>

Query Match 31.7%; Score 53.5; DB 2; Length 159;
 Best Local Similarity 44.4%; Pred. No. 7.2;
 Matches 12; Conservative 3; Mismatches 9; Indels 3; Gaps 1;
 Db 6 6 GAARGRASCW---ALALLMLAVPGW 29
 10 GSSRSRSCWCKNLKQLAFGLADDPAM 36

RESULT 6
 A82729
 Phosphatidate cytidyltransferase XF1049 [imported] - *Xylella fastidiosa* (strain 955c)
 C:Species: *Xylella fastidiosa*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: A82729
 R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequence
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: A82729
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-284 <SIM>
 A:Cross-references: GB:AE003942; GB:AE003849; NID:99105990; PIDN:AAF83859.1; GSPDB:GN0012
 A:Experimental source: strain 955c
 A:Authors: A.J.G. Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 rtiens, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Caniaro, L.E.A.; Carraro, D.M.; Carre, H
 S-Neto, E.; Docena, C.; El-Dorzy, H.; Facincani, A.P.; Ferreira, A.U.S.
 Submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
 U.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A90879; MUID:86218082; PMID:3754790
A:Accession: A01398
A:Molecule type: DNA
A:Residues: 1-14 <CAI>
A:Experimental source: newborn calf testis, clones cblm15 and ps21
A:Accession: B01398
A:Molecule type: mRNA
A:Residues: 15-575 <CA2>
C:Comment: This glycoprotein, produced by the Sertoli cells of the testis, causes regression of Mullerian duct origin. Other roles for this protein in gonadal differentiation ter duct regression and in the adult ovary.
C:Comment: This protein is homologous to the beta transforming growth factor, inhibin A, these sequences. All of these proteins are biologically active as disulfide-linked dimers
C:Comment: Although it does not compete with EGF for receptor binding sites, MIS can inhibit C:Superfamily: inhibin
C:Keywords: cytochrome; glycoprotein; gonadal differentiation; testis
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-24/Domain: propeptide #status predicted <PRO>
F:25-575/Product: mullerian inhibiting factor #status predicted <MAT>
F:18,344/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match
Best Local Similarity 31.1%; Score 52.5; DB 1; Length 575;
Matches 11; Conservative 4; Mismatches 7; Indels 3; Gaps 1;

Db
4 GAGAAAGRAAGCAALALAVVP 28
41 GAGALIFQADMPUSLWL---PG 62

RESULT 12
A45054
Probable intercellular signal transducer or transmitter Fz-1 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 02-Mar-2001
C:Accession: A45054
R:Chan, S.D.; Karpf, D.B.; Fowlkes, M.E.; Hooks, M.; Bradley, M.S.; Vuong, V.; Bambino, J. Biol. Chem. 267, 25202-25207, 1992
A:Title: Two homologs of the Drosophila polarity gene frizzled (Fz) are widely expressed
A:Reference number: A45054; MUID:93094228; PMID:1334084
A:Accession: A45054
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-641 <CHAS>
A:Experimental source: UMR 106 osteosarcoma cell line
A:Note: Sequence extracted from NCBI Backbone (NCBIP:120154)
C:Superfamily: fruit fly frizzled protein

Query Match
Best Local Similarity 30.8%; Score 52; DB 2; Length 641;
Matches 15; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

Db
5 AGAAGRA-SWCNA---LALALAVVP 27
38 AGHRPRRAHRCWARGLLILMLLEAP 64

RESULT 13
S30075
Fertic reductase (EC 1.6.99.-) FRE1 - Yeast (Saccharomyces cerevisiae)
N:Alternate names: protein L8167.2; protein YLR214W
C:Species: Saccharomyces cerevisiae
C>Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 17-Mar-2000
C:Accession: S30075; S48565
R:Dancs, A.; Roman, D.G.; Anderson, G.U.; Hinebush, A.G.; Klausner, R.D.
Proc. Natl. Acad. Sci. U.S.A. 89, 3869-3873, 1992
A:Title: Fertic reductase of Saccharomyces cerevisiae: Molecular characterization, role
A:Reference number: S30075; MUID:92237270; PMID:1570306
A:Accession: S30075
A:Molecule type: DNA
A:Residues: 1-686 <DAN>
A:Cross-references: EMBL:M86908; NID:g171520; PIDN:AAA34608.1; PID:g171521
R:Pauley, A.

submitted to the EMBL Data Library, September 1994
A:Description: The sequence of S. cerevisiae cosmid 8167.
A:Reference number: S48545
A:Accession: S48565
A:Molecule type: DNA
A:Residues: 1-686 <PAU>
A:Cross-references: EMBL:U14913; NID:g544497; PIDN:AAB67424.1; PID:g544499; MIPS:YLR214W
C:Genetic: SGD:PRE1
A:Cross-references: SGD:S0004204; MIPS:YLR214W
A:Map position: 12R
C:Superfamily: ferric reductase PRE2
C:Keywords: oxidoreductase; transmembrane protein
F:6-22/Domain: transmembrane #status predicted <TM1>
F:152-168/Domain: transmembrane #status predicted <TM2>
F:216-232/Domain: transmembrane #status predicted <TM3>
F:258-274/Domain: transmembrane #status predicted <TM4>
F:298-314/Domain: transmembrane #status predicted <TM5>
F:329-345/Domain: transmembrane #status predicted <TM6>
F:359-375/Domain: transmembrane #status predicted <TM7>
F:531-547/Domain: transmembrane #status predicted <TM8>

Query Match
Best Local Similarity 30.5%; Score 51.5; DB 2; Length 686;
Matches 8; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

Db
11 RASNC-WALALMLAVP 26
144 RSQWCAGLVFVFWAVL 160

RESULT 14
A38587
collagen, cornea-specific - chicken
C:Species: Gallus gallus (chicken)
C>Date: 14-Feb-1992 #sequence_revision 15-Aug-1997 #text_change 15-Sep-2003
C:Accession: S16501; A38587
R:Marchant, J.K.; Linsemayer, T.F.; Gordon, M.K.
Proc. Natl. Acad. Sci. U.S.A. 89, 1560-1564, 1991
A:Title: cDNA analysis predicts a cornea-specific collagen.
A:Reference number: A38587; MUID:91142213; PMID:1705041
A:Accession: S16501
A:Molecule type: mRNA
A:Residues: 1-1146 <MAR>
A:Cross-references: EMBL:M60172; NID:g211609; PIDN:AAA48703.1; PID:g211610
A:Accession: A38587
A:Molecule type: mRNA
A:Residues: 1-174, 'X', 176-233, 'X', 235-344, 'X', 346-408, 'X', 410-499, 'X', 501-876, 'X', 878-114
A:Cross-references: GB:M60172
C:Keywords: cornea

Query Match
Best Local Similarity 30.5%; Score 51.5; DB 2; Length 1146;
Matches 15; Conservative 2; Mismatches 6; Indels 13; Gaps 3;

Db
4 GAGAAAGRA-SWCNA---LALALAVVP 26
29 GAGALGAPSPWCPGSCGSMWKLGLLMLL 64

RESULT 15
E95984
conserved hypothetical membrane protein, paralogue of Y21292 SMB20848 [imported] - Sinori
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95984
R:Finan, T.M.; Weidner, S.; Wong, K.; Bhurestter, J.; Chain, P.; Vorholter, F.U.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95984
A:Status: preliminary
A:Molecule type: DNA

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

3M protein - protein search, using sw model

Run on: June 8, 2004, 13:56:38 ; Search time 1.25 Seconds

(without alignments)
1249.684 Million cell updates/sec

Title: US-10-010-050a-2_COPY_1_30
Perfect score: 169

Sequence: 1 MRGAGAGRGASGCMWALALWLAVDPGWS 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	169	100.0	407	1	CLNS_HUMAN
2	56.5	33.4	126	1	CRGB_RALSO
3	53.5	31.7	162	1	HXA9_CAVPO
4	53	31.4	398	1	GSP_KLEPN
5	52.5	31.1	575	1	MIS_BOVIN
6	52	30.8	374	1	OPSI_DROPS
7	52	30.8	641	1	FZDI_RAT
8	51.5	30.5	686	1	FRRI_YEAST
9	50	29.6	106	1	SH_RAT
10	50	29.6	219	1	EREL_YEAST
11	49.5	29.3	514	1	APPC_ECOLI
12	49.5	29.3	92	1	DEP3_HUMAN
13	49.5	29.3	330	1	PEXD_CAEEL
14	49.5	29.3	519	1	INT_SYNY3
15	49	29.0	286	1	CLNG_HUMAN
16	48	28.4	221	1	SDPL_HUMAN
17	48	28.4	233	1	PTCL_AGRYS
18	48	28.4	718	1	SLU3_BOVIN
19	48	28.4	718	1	SLU3_CANFA
20	48	28.4	718	1	SLU3_HUMAN
21	48	28.4	718	1	SI53_MOUSE
22	48	28.4	1245	1	NIDO_MOUSE
23	47.5	28.1	599	1	YV06_CAEEL
24	47.5	28.1	640	1	CAN5_HUMAN
25	47.5	28.1	640	1	CAN5_MOUSE
26	47.5	28.1	640	1	CAN5_MOUSE
27	47	27.8	339	1	VIRG_ECOLI
28	47	27.8	343	1	GAS1_MOUSE
29	47	27.8	345	1	GAS1_MOUSE
30	47	27.8	508	1	Y202_HUMAN
31	46.5	27.5	521	1	CYOA_HAEN
32	46.5	27.5	332	1	YIAX_HAEN
33	46.5	27.5	373	1	OPSL_DROME

34	46	27.2	108	1	YNFA_SALTI	Q86622 salmonella
35	46	27.2	335	1	XTNB_STRLI	P26515 streptomyc
36	46	27.2	812	1	U84A_HUMAN	O94901 homo sapien
37	45.5	26.9	108	1	YNFA_SALTY	O88p52 salmonella
38	45.5	26.9	249	1	MSX1_CHICK	P26361 gallus gall
39	45.5	26.9	420	1	MLIC_XENLA	P49219 xenopus lae
40	45.5	26.9	462	1	NARU_ECOLI	P37758 escherichia
41	45.5	26.9	471	1	MELB_ENTAE	O07366 enterobacte
42	45.5	26.9	471	1	MELB_KLEPN	O02581 klebsiella
43	45.5	26.9	844	1	PPPB_ECOLI	P02919 e penicillii
44	45.5	26.9	1112	1	CN3B_HUMAN	Q13370 homo sapien
45	45.5	26.9	2832	1	NDVB_RHIME	P20471 rhizobium m

ALIGNMENTS

RESULT 1	ID	CLNS_HUMAN	STANDARD	PRT	407 AA.
AC	078503				
DT	15-DEC-1998	(Rel. 37, Created)			
DT	15-DEC-1998	(Rel. 37, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	Ceroid-lipofuscinosis neuronal protein 5 (CLNS protein).				
GN	CLNS.				
OS	Homo sapiens (human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A., VARIANT VLINCL ASN-279, AND VARIANT ARG-368.				
RC	TISSUE=Fetal brain;				
RC	MEDLINE=98324783; PubMed=9662406;				
RA	Savukoski M., Klockars T., Holmberg V., Santavuori P., Lander E.S.,				
RA	Peltonen L.;				
RT	"CLNS, a novel gene encoding a putative transmembrane protein mutated				
RT	in Finnish variant late infantile neuronal ceroid lipofuscinosis.",				
RL	Nat. Genet. 19:286-288(1998).				
RN	[2]				
RP	SUBCELLULAR LOCATION AND GLYCOSYLATION.				
RX	MEDLINE=21968572; PubMed=11971870;				
RA	Isoompi J., Vesa J., Jalanko A., Peltonen L.;				
RT	"Lysosomal localization of the neuronal ceroid lipofuscinosis CLNS				
RT	protein.",				
RL	Hum. Mol. Genet. 11:885-891(2002).				
CC	-1- FUNCTION: Not known.				
CC	-1- SUBCELLULAR LOCATION: Lysosomal.				
CC	-1- TISSUE SPECIFICITY: Ubiquitous.				
CC	-1- PTM: Glycosylated.				
CC	-1- DISEASE: Defects in CLNS are the cause of Finnish variant late-				
CC	infantile neuronal ceroid lipofuscinosis (VLINCL) [MIM:256711];				
CC	also known as ceroid lipofuscinosis neuronal 5 (CLNS). VLINCL is a				
CC	fatal childhood neurodegenerative disease characterized by				
CC	progressive visual and mental decline, motor disturbance, epilepsy				
CC	and behavioral changes. The first symptom is motor clumsiness,				
CC	followed by progressive visual failure, mental and motor				
CC	deterioration and later by myoclonia and seizures.				
CC	-1- DATABASE: NAME=NCL CLNS;				
CC	NOTE=Neural Ceroid Lipofuscinoses mutation db;				
CC	WWW="http://www.ucl.ac.uk/ncl/CLNS.html".				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.ebi.ac.uk/announcements/				
CC	or send an email to license@ebi.ac.uk).				
CC	-----				
DR	EMBL; AF068227; AAC27614.1; -.				
DR	GeneW; HGNC:2076; CLNS.				
DR	MIT; 608102; -.				

DR MIM; 256731; -.
DR GO; GO:0016021; C: integral to membrane; TAS.
DR GO; GO:0008151; P: cell growth and/or maintenance; TAS.
KW Transmembrane; Lysosome; Glycoprotein; Neuronal ceroid lipofuscinosis;
KW Disease mutation; Polymorphism; Epilepsy.
FT TRANSMEM 75 91 POTENTIAL
FT CAROHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CAROHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT VARIANT 279 279 D -> N (in VLINCL) .
FT VARIANT 368 368 /FTID-VAR 005137.
FT VARIANT 46339 MW; 449702DIDC9BFE54 CRC64; /FTID-VAR 005138.
FT SEQUENCE 407 AA; 46339 MW; 449702DIDC9BFE54 CRC64;
Query Match 100.0%; Score 169; DB 1; Length 407;
Best Local Similarity 100.0%; Pred. No. 1,4e-14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0
2y 1 MRGGAARGASWCMALALLMLAVVPGWS 30
Db 62 MRGGAARGASWCMALALLMLAVVPGWS 91

RESULT 2
CRCB_PALSO
ID ID CRCB_PALSO STANDARD, PRT, 126 AA.
KC Q8XZR2;
YT 28-FEB-2003 (Rel. 41, Created)
YT 28-FEB-2003 (Rel. 41, Last sequence update)
YT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein crcb homolog.
IN CRCB OR RSC133 OR RS02855.
NS Ralstonia solanacearum (Pseudomonas solanacearum).
NC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
NC Burkholderiaceae; Ralstonia.
NC NCBI_TaxID=305;
[1]
SEQUENCE FROM N.A.
ID STRAIN=GMT1000;
MEDLINE=21681879; PubMed=11823852;
LA Salanoubat M., Genin S., Artiguenave F., Gonzy J., Mengerot S.,
LA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
LA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
LA Gaspin C., Lavie M., Molan A., Robert C., Saurin W., Schiek T.,
LA Signier P., Thebaud P., Whalen M., Wincker P., Levy M.,
LA Weissenbach J., Boucher C.A.;
JT "Genome sequence of the plant pathogen Ralstonia solanacearum.",
JT Nature 415:497-502 (2002).
IL -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
-1- SIMILARITY: Belongs to the crcb family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL AL646064, CAD15035.1; -.
HAMAP; MF_00454; -. 1.
InterPro; IPR003691; Camphor_CrcB.
Pfam; PF02537; CRCB, 1.
TIGRFAMs; TIGR00494; crcb, 1.
Transmembrane; Complete proteome.
TRANSMEM 5 24 POTENTIAL.
TRANSMEM 34 56 POTENTIAL.

```

FT TRANSMEM 68 90 POTENTIAL.
FT TRANSMEM 100 122 POTENTIAL.
SQ SEQUENCE 126 AA; 13101 MW; C2443FBAE5C81CB3 CRC64;

Query Match
Best Local Similarity 33.4%; Score 56.5; DB 1; Length 126;
Matches 13; Conservative 3; Mismatches 6; Indels 3; Gaps 2

Dy 4 GAGAGRGRASWC-WALALLWLAVP 27
||| | | | | | | | | | | | | | |
10 GVGALLG-AMLRWAFVILWNLINP 32

RESULT 3
HXA9_CAVPO
ID HXA9_CAVPO STANDARD; PRT; 162 AA.
AC P51783;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein Hox-A9 (Hox-1.7) (Fragment).
GN HoxA9 OR Hox-1.7.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathii; Cavidae; Cavia.
OC NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=92190539; PubMed=1963703;
RA Rubin M.R.; Nguyen-Huu M.C.;
RT "Alternatively spliced Hox-1.7 transcripts encode different protein
products."
RL DNA Seq. 1115-124(1190).
CC -1- FUNCTION: Sequence-specific transcription factor which is part of
a developmental regulatory system that provides cells with
specific positional identities on the anterior-posterior axis.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the Abd-B homeobox family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; X13536; CAA31887.1; -.
DR HSSP; P02833; NANT.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR006711; Hox9 act.
DR InterPro; IPR000047; HTH_1ambrepresr.
DR Pfam; PF00046; homeobox; 1.
DR Pfam; PF04617; Hox9_act; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
FT Transcription regulation.
FT NON_TER 1
FT DNA_BIND 96 155 HOMEBOX.
SQ SEQUENCE 162 AA; 18661 MW; B84FA507EC67CF91 CRC64;

Query Match
Best Local Similarity 31.7%; Score 53.5; DB 1; Length 162;
Matches 12; Conservative 3; Mismatches 9; Indels 3; Gaps 1;

```

2b 13 GSSRRSSMCNKLQLAFAAGLADPAW 39

RESULT 4

ISPL_KLEPN STANDARD; PRT; 398 AA.

15751;

01-APR-1990 (Rel. 14, Created)

01-APR-1990 (Rel. 14, Last sequence update)

01-NOV-1995 (Rel. 32, Last annotation update)

General secretion pathway protein L (Pululanase secretion protein)

PUL.

Klebsiella pneumoniae.

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Klebsiella.

NCBI_TaxID=573;

[1]

SEQUENCE FROM N.A.

SPRAIN-UNF 5023;

MEDLINE=90286914; PubMed=2162463;

Pugsley A.P., Reys I.,

"Five genes at the 3' end of the Klebsiella pneumoniae pulc operon

are required for pululanase secretion."

Mol. Microbiol. 4:365-379(1990).

-1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE

EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF PULULANASE.

-1- SIMILARITY: BELONGS TO THE EXEL/PULU/OUTL/XCPY FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; M32613; AAA25134.1; -

EMBL; X52462; CAA36697.1; -

PIR; S11800; S11800.

InterPro; IPR007812; GspL.

Pfam; PF05134; GspL; 1.

Transport; Transmembrane; Inner membrane.

DOMAIN 224 248 ARG-RICH.

TRANSMEM 249 265. POTENTIAL.

SEQUENCE 398 AA; 44198 MW; 72CF5D87DBA99948 CRC64;

Query Match 31.4%; Score 53; DB 1; Length 398;

Best Local Similarity 35.1%; Pred. No. 9.9;

Matches 13; Conservative 3; Mismatches 11; Indels 10; Gaps 1;

4 GAGAGRGRSMCWALALMLAVP-----GMS 30

118 GGGKRRAMQWCDLIGLVLAITPVLLALPHSPTGMS 154

4 GAGAGRGRSMCWALALMLAVP-----GMS 30

118 GGGKRRAMQWCDLIGLVLAITPVLLALPHSPTGMS 154

4 GAGAGRGRSMCWALALMLAVP-----GMS 30

118 GGGKRRAMQWCDLIGLVLAITPVLLALPHSPTGMS 154

4 GAGAGRGRSMCWALALMLAVP-----GMS 30

118 GGGKRRAMQWCDLIGLVLAITPVLLALPHSPTGMS 154

4 GAGAGRGRSMCWALALMLAVP-----GMS 30

118 GGGKRRAMQWCDLIGLVLAITPVLLALPHSPTGMS 154

4 GAGAGRGRSMCWALALMLAVP-----GMS 30

118 GGGKRRAMQWCDLIGLVLAITPVLLALPHSPTGMS 154

4 GAGAGRGRSMCWALALMLAVP-----GMS 30

118 GGGKRRAMQWCDLIGLVLAITPVLLALPHSPTGMS 154

4 GAGAGRGRSMCWALALMLAVP-----GMS 30

118 GGGKRRAMQWCDLIGLVLAITPVLLALPHSPTGMS 154

SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RA Cate R.L., Mattaliano R.J., Hession C., Tizard R., Farber N.M.,

RA Cheung A., Ninta E.G., Frey A.Z., Gash D.J., Chow E.P., Fisher R.A.,

RA Bertonis J.M., Torres G., Wallner B.P., Ramachandran K.L.,

RA Ragin R.C., Manganaro T.F., McLaughlin D.T., Donahoe P.K.,

RT "Isolation of the bovine and human genes for Mullerian inhibiting

substance and expression of the human gene in animal cells";

RL Cell 45:685-698(1986).

-1- FUNCTION: THIS GLYCOPROTEIN, PRODUCED BY THE SEROLI CELLS OF THE

TESTIS, CAUSES REGRESSION OF THE MULLERIAN DUCT. IT ALSO IS ABLE

TO INHIBIT THE GROWTH OF TUMORS DERIVED FROM TISSUES OF MULLERIAN

DUCT ORIGIN.

-1- SUBUNIT: Homodimer; disulfide-linked.

-1- MISCELLANEOUS: Although it does not compete with EGF for receptor

binding sites, MIS can inhibit the autophosphorylation of the EGF

receptor in vitro.

-1- SIMILARITY: Belongs to the TGF-beta family.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; M3151; AAA98765.1; -

PIR; A01398; MFBOM.

InterPro; IPR006799; AMH N.

InterPro; IPR002400; GF Cysnot.

InterPro; IPR001839; TGFb.

Pfam; PF04709; AMH N; 1.

Pfam; PF00019; TGF-beta; 1.

PRINTS; PR00438; GFCYSNOT.

PRODOM; PD000357; TGFb; 1.

SMART; SM00204; TGFb; 1.

PROSITE; PS00250; TGF-beta; 1.

Growth factor; Glycoprotein; Gonadal differentiation; Signal.

FT SIGNAL 1 17

FT PROPEP 18 24

FT CHAIN 25 575

FT DISULFID 477 541

FT DISULFID 503 572

FT DISULFID 507 574

FT DISULFID 540 540

FT CARBOHYD 78 78

FT CARBOHYD 344 344

SEQUENCE 575 AA; 60623 MW; 892B89C11ACB85A8 CRC64;

Query Match 31.1%; Score 52.5; DB 1; Length 575;

Best Local Similarity 44.0%; Pred. No. 16;

Matches 11; Conservative 4; Mismatches 7; Indels 3; Gaps 1;

4 GAGAGRGRSMCWALALMLAVPG 28

41 GSGALIFQQAQWDPWSLWL--PG 62

4 GAGAGRGRSMCWALALMLAVPG 28

41 GSGALIFQQAQWDPWSLWL--PG 62

4 GAGAGRGRSMCWALALMLAVPG 28

41 GSGALIFQQAQWDPWSLWL--PG 62

4 GAGAGRGRSMCWALALMLAVPG 28

41 GSGALIFQQAQWDPWSLWL--PG 62

4 GAGAGRGRSMCWALALMLAVPG 28

41 GSGALIFQQAQWDPWSLWL--PG 62

4 GAGAGRGRSMCWALALMLAVPG 28

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Apple Hill;
RX MEDLINE=93012921; PubMed=1398053;
RA Carnali J.P., Harli D.L.;
RT "Variable rates of evolution among Drosophila opsin genes.";
RL Genetics 132:193-204(1992).
CC -!- FUNCTION: Visual pigments are the light-absorbing molecules that
CC mediate vision. They consist of an apoprotein, opsin, covalently
CC linked to cis-retinal.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- PTM: Some or all of the carboxyl-terminal Ser or Thr residues may
CC be phosphorylated.
CC -!- MISCELLANEOUS: Each Drosophila eye is composed of 800 facets or
CC ommatidia. Each ommatidium contains 8 photoreceptor cells (R1-R8),
CC the R1 to R6 cells are outer cells, while R7 and R8 are inner
CC cells.
CC -!- MISCELLANEOUS: Opsin Rh1 has an absorption maximum at 480 nm.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC Opsin subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
JR EMBL: X65877; CAA46708.1; -.
JR PIR: S40691; S40691.
JR HSP: P02699; 1RDV.
JR FlyBase: FBgn012733; Dpse\unnae.
JR InterPro: IPR000276; GPCR_Rhodpsn.
JR Pfam: PR00001; 7tm.1; 1.
JR PRINTS: PS00237; GPCRHOPOPSN.
JR PROSITE: PS00262; G_PROTEIN_RECEP_F1_1; 1.
JR PROSITE: PS00238; OPSIN; 1.
JR GlycoProtein: Retinal protein; Transmembrane; Phosphorylation;
JR Glycoprotein; G-protein coupled receptor; Vision.
JR DOMAIN 1 49
JR TRANSMEM 50 74
JR DOMAIN 75 86
JR TRANSMEM 87 112
JR DOMAIN 113 126
JR TRANSMEM 127 146
JR DOMAIN 147 165
JR TRANSMEM 166 190
JR DOMAIN 191 214
JR TRANSMEM 215 242
JR TRANSMEM 243 277
JR TRANSMEM 278 301
JR DOMAIN 302 308
JR TRANSMEM 309 333
JR DOMAIN 334 374
JR CARBOHYD 20 20
JR DISULFID 123 201
JR BINDING 320 320
JR SEQUENCE 374 AA; 41604 MW; 8228082970F34 CRC64;

Query Match 30.8%; Score 52; DB 1; Length 374;
Best Local Similarity 32.0%; Pred. No. 13;
Matches 8; Conservative 7; Mismatches 8; Indels 2; Gaps 1;
Y 8 ARGRASMCWALMTMLAVP--GMS 30
b 166 ALGKIAIVPMFTWCCLAPEFGMS 190
RESULT 7
ZDI_RAT

```

```

ID FZD1 RAT STANDARD: PRT; 641 AA.
AC 008463;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Frizzled 1 precursor (Frizzled-1) (Fz-1) (Fz1).
GN FZD1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; Tissue=Osteosarcoma;
RX MEDLINE=9304228; PubMed=1334084;
RA Chan S.D.H., Karpf D.B., Fowlkes M.E., Hooks M., Bradley M.S.,
RA Vuong V., Barbino T., Liu M.Y.C., Arnaut C.D., Strewler G.J.,
RA Nissen R.A.;
RT "Two homologs of the Drosophila polarity gene frizzled (fz) are widely
RT expressed in mammalian tissues."
RL J. Biol. Chem. 267:25202-25207(1992).
RN [2]
RP COUPLING TO BETA-CATENIN PATHWAY.
RX MEDLINE=9324245; PubMed=1039542;
RA Sheldahl L.C., Park M., Malbon C.C., Moon R.T.;
RT "Protein kinase C is differentially stimulated by Wnt and Frizzled
RT homologs in a G-protein-dependent manner."
RL Curr. Biol. 9:695-698(1999).
CC -!- FUNCTION: Receptor for Wnt proteins. Most of frizzled receptors
CC are coupled to the beta-catenin canonical signaling pathway, which
CC leads to the activation of dishevelled proteins, inhibition of
CC GSK-3 kinase, nuclear accumulation of beta-catenin and activation
CC of Wnt target genes. A second signaling pathway involving PKC and
CC calcium fluxes has been seen for some family members, but it is
CC not yet clear if it represents a distinct pathway or if it can be
CC integrated in the canonical pathway, as PKC seems to be required
CC for Wnt-mediated inactivation of GSK-3 kinase. Both pathways seem
CC to involve interactions with G-proteins. May be involved in
CC transduction and intercellular transmission of polarity
CC information during tissue morphogenesis and/or in differentiated
CC tissues. Activation by Wnt8 induces expression of beta-catenin
CC target genes.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Widely expressed. Most abundant in kidney,
CC liver, uterus, ovary and heart. Lower levels seen in brain and
CC intestine. Extremely low in calvaria, mammary glands and testis.
CC -!- DEVELOPMENTAL STAGE: Expressed predominantly in neonatal tissues,
CC at lower levels in adult.
CC -!- DOMAIN: Lys-Thr-X-X-X-Tyr motif is involved in the activation of
CC the Wnt/beta-catenin signaling pathway (by similarity).
CC -!- DOMAIN: The Fz domain is involved in binding with Wnt ligands (by
CC similarity).
CC -!- SIMILARITY: Belongs to the Fz/Smo G-protein coupled receptor
CC family.
CC -!- SIMILARITY: Contains 1 frizzled (Fz) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
JR EMBL: L02529; AAA1173.1; -.
JR InterPro: IPR000539; Frizzled.
JR InterPro: IPR000024; Fz domain.
JR Pfam: PR01534; Fz domain.
JR PRINTS: PR01392; Fz_1.
JR PRINTS: PR00469; FRIZZLED.
JR SMART: SM00063; FRI; 1.
JR PROSITE: PS00308; FZ; 1.

```

```

DR PROSITE; P550261; G PROTEIN RECEPTOR; Transmembrane;
KW Multigene family; G-protein coupled receptor; Transmembrane;
KW Developmental protein; Wnt signaling pathway; Glycoprotein; Signal.
FT SIGNAL 1 68
FT CHAIN 69 641
FT DOMAIN 69 316 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 317 337 1 (POTENTIAL).
FT TRANSMEM 338 348 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 349 369 2 (POTENTIAL).
FT DOMAIN 370 396 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 397 417 3 (POTENTIAL).
FT TRANSMEM 418 439 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 440 460 4 (POTENTIAL).
FT TRANSMEM 461 483 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 484 504 5 (POTENTIAL).
FT TRANSMEM 505 530 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 531 551 6 (POTENTIAL).
FT TRANSMEM 552 595 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 596 616 7 (POTENTIAL).
FT DOMAIN 617 641 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 106 224 F2.
FT SITE 85 90 POLY-PRO.
FT SITE 619 624 LYS-THR-X-X-TRP MOTIF.
FT CARBOHYD 125 125 PDZ-BINDING.
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 641 AA; 71027 MW; D82E2C13E81B8B6 CRC64;

Query Match 30.5%; Score 52; DB 1; Length 641;
Best Local Similarity 55.6%; Pred. No. 20;
Matches 15; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

Y 5 AGAARGRA-SWCA--LALLMLAVVP 27
b 38 AGHRRPRHRCVARGLLMLLEAP 64

RESULT 8
REL_YEAST STANDARD; PRT; 686 AA.
D FREI_YEAST
C P32791;
T 01-OCT-1993 (Rel. 27, Created)
T 01-OCT-1993 (Rel. 27, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E Ferric reductase transmembrane component 1 precursor (BC 1.16.1.7)
E (Ferric-chelate reductase 1).
FREL OR YLR1214W OR L8167.2.
N Saccharomyces cerevisiae (Baker's yeast).
C Saccharomycetes;
C Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
C Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
C NCBI_TaxID=4932;
C (1)
P SEQUENCE FROM N.A.
K STRAIN=F113;
K MEDLINE=92237270; PubMed=1570306;
K Dancis A., Roman D.G., Anderson G.J., Hinnebusch A.G., Klausner R.D.;
K "Ferric reductase of Saccharomyces cerevisiae: molecular
K characterization, role in iron uptake, and transcriptional control by
K iron.";
K Proc. Natl. Acad. Sci. U.S.A. 89:3869-3873(1992).

SEQUENCE FROM N.A.
STRAIN=S288C / AB972;
MEDLINE=9713267; PubMed=9169871;
J Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansoerge W.,
J Benes V., Bruckner M., Delius H., Dubois E., Duesterhoelt A.,
J Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koelter P.,
J Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
J Mueller-Auer S., Neutwich U., Obermaier B., Piravandi E., Pohl T.M.,
J Portetelle D., Purnelle B., Rechner S., Rieger M., Rinke M., Rose M.,
J Scharfe M., Scherens B., Scholler P., Schwaiger C., Schwarz S.,
J Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,

```

```

RA Viereendeels F., Voet M., Volckaert G., Voss H., Wandurt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hant J., Hohnsels J.D.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RA Nature 387:87-90(1997).
RA (3)
RP CHARACTERIZATION.
RP MEDLINE=93057491; PubMed=1431884;
RA Anderson G.J., Lesuisse E., Dancis A., Roman D.G., Labbe P.,
RA Klausner R.D.;
RA "Ferric iron reduction and iron assimilation in Saccharomyces
RA cerevisiae.";
RL J. Inorg. Biochem. 47:249-255(1992).
CC -1- FUNCTION: Reductase activity that acts on ferric iron chelates
CC external to the cell. Plays a role in iron uptake. May
CC participate in the transport of electrons from cytoplasm to an
CC extracellular substrate (ferric ion) via FAD and heme
CC intermediates. May also participate in Cu(II) reduction and Cu(I)
CC uptake.
CC -1- CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).
CC -1- COFACTOR: FAD (Probable).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- INDUCTION: BY IRON DEPRIVATION. REPRRESSED BY IRON UPTAKE.
CC -1- SIMILARITY: Belongs to the FRE / CYBB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M66908; AAA34608.1; -
CC EMBL; U14913; AAB67424.1; -
CC PIR; S30075; S30075.
CC Germonline; 142276; -
CC DR SGD; S0004204; FRE1.
CC DR GO; GO:0005886; C:plasma membrane; IDA.
CC DR GO; GO:0000293; F:ferric-chelate reductase activity; IDA.
CC DR GO; GO:0015677; P:copper ion import; IDA.
CC DR GO; GO:0006826; P:iron ion transport; IDA.
CC DR InterPro; IPR002916; Ferric_reduct.
CC DR Pfam; PF01794; Ferric_reduct.
CC KW Oxidoreductase, Electron transport; Transmembrane; Iron transport;
CC FAD; NAD; copper; Glycoprotein; Signal; Multigene family.
CC FT SIGNAL 1 22
CC CHAIN 23 686
CC FT NP BIND 462 468 FAD (POTENTIAL).
CC FT NP BIND 532 540 NAD (POTENTIAL).
CC FT TRANSMEM 147 169 1 (POTENTIAL).
CC FT TRANSMEM 216 236 2 (POTENTIAL).
CC FT TRANSMEM 258 277 3 (POTENTIAL).
CC FT TRANSMEM 296 316 4 (POTENTIAL).
CC FT TRANSMEM 329 348 5 (POTENTIAL).
CC FT TRANSMEM 369 397 6 (POTENTIAL).
CC FT TRANSMEM 529 550 7 (POTENTIAL).
CC FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 686 AA; 78853 MW; 7F6BB3B9A95D6A3 CRC64;

Query Match 30.5%; Score 51.5; DB 1; Length 686;
Best Local Similarity 47.1%; Pred. No. 25;
Matches 8; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 11 RASWC-WALLMLAVVP 26
Db 144 RSCWCAMGLVFFWAVL 160

```

```

RESULT 9
SH_RAT
ID SH_RAT STANDARD; PRT; 106 AA.
AC P55248;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Putative protein SH (SH-4).
DS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC NCBI_TaxID=10116;
CC [1]
CC SEQUENCE FROM N.A.
CC TISSUE=Heart;
CC MEDLINE=87149087; PubMed=3547652;
CC Adelman J.P., Bond C.T., Douglass J., Herbert E.;
CC "Two mammalian genes transcribed from opposite strands of the same
CC DNA locus.";
CC Science 235:1514-1517 (1987).
CC -1- FUNCTION: May be involved with the regulation of GNRH gene
CC expression. It is not known if this protein is transcribed.
CC -1- TISSUE SPECIFICITY: Heart.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M15527; AAA42140.1; -.
CC PIR; A29760; A29760.
CC SEQUENCE 106 AA; 11792 MW; F9BF8FDBFEAABR3E CRC64;

Query Match 29.6%; Score 50; DB 1; Length 106;
Best Local Similarity 47.1%; Pred. No. 7.5;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Y 5 AGAAGGASGAWMALAL 21
D 2 AHAVRSKSNWCQTYLL 18

RESULT 10
RP1_YEAST
ID RP1_YEAST STANDARD; PRT; 219 AA.
AC Q05359;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ERP1 protein precursor.
CC ERP1 OR YAR002C-A OR YAR002A.
CC Saccharomyces cerevisiae (Baker's yeast).
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
CC NCBI_TaxID=4932;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=S288c / AB972;
CC MEDLINE=95028152; PubMed=7941740;
CC Clark M.W., Keng T., Storms R.K., Zhong W., Fortin N., Zeng B.,
CC Delaney S., Ouellette B.F.F., Barton A.B., Kaback D.B., Bussey H.;
CC "Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of
CC the 42 kbp SPO7-CENT-CDC15 region.";
CC Yeast 10:535-541 (1994).
CC [2]
CC CHARACTERIZATION.
CC MEDLINE=99287733; PubMed=10359606;
CC Marzloch M., Henthorn D.C., Herrmann J.M., Wilson R., Thomas D.Y.,
CC Bergeron J.J., Solari R.C., Rowley A.;

```

```

RT "Erp1p and Erp2p, partners for Emp24p and Erv25p in a yeast p24
RT complex.";
RT Mol. Biol. Cell 10:1923-1938 (1999).
CC -1- FUNCTION: Involved in vesicular protein trafficking.
CC -1- SUBUNIT: Associates with EMP24, ERV25 and ERP2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Endoplasmic
CC reticulum (By similarity).
CC -1- SIMILARITY: Belongs to the EMP24/GP25L family.
CC -1- SIMILARITY: Contains 1 GOLD domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L22015; AAC04958.1; -.
CC GenomOnline; 140222; -.
CC SGP; S0002129; ERP1.
CC GO; GO:0030138; C:COPII-coated vesicle; IDA.
CC GO; GO:0006888; P:ER to Golgi transport; IPI.
CC InterPro; IPR000348; Emp24_SP25L_p24.
CC InterPro; IPR009038; GOLD.
CC Pfam; PF01105; EMP24_GP25L; 1.
CC PROSITE; PSS0866; GOLD; 1.
CC Transports; Protein transport; Transmembrane; Signal;
CC Endoplasmic reticulum.
CC SIGNAL 1 22
CC CHAIN 23 219
CC DOMAIN 23 186
CC TRANSMEM 187 207
CC DOMAIN 208 219
CC FT DOMAIN 32 131
CC FT SEQUENCE 219 AA; 24723 MW; 16236C1886FC42E8 CRC64;

Query Match 29.6%; Score 50; DB 1; Length 219;
Best Local Similarity 39.3%; Pred. No. 14;
Matches 11; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 2 PRGAGAGRASGAWMALALMLAVPGM 29
D 177 RDASEAIVNSRAMWVIYQLIVAVTCGM 204

RESULT 11
APPC_ECOLI
ID APPC_ECOLI STANDARD; PRT; 514 AA.
AC P26459;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome BD-II oxidase subunit I (EC 1.10.3.-).
CC APPC OR CYXA OR CBDA OR B0978.
CC Escherichia coli.
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Escherichia.
CC NCBI_TaxID=562;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=K12;
CC MEDLINE=92049231; PubMed=1658595;
CC Dassa J., Feihl H., Marck C., Dion M., Kieffer-Bontemps M.,
CC Boquet P.L.;
CC "A new oxygen-regulated operon in Escherichia coli comprises the
CC genes for a putative third cytochrome oxidase and for pH 2.5 acid
CC RT phosphatase (appa).";
CC Mol. Gen. Genet. 229:341-352 (1991).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=K12 / MG1655;
CC MEDLINE=97426617; PubMed=9278503;

```

Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

"The complete genome sequence of *Escherichia coli* K-12.";

Science 277:1453-1474(1997).

[3]

SEQUENCE FROM N.A.

STRAIN=K12;

MEDLINE=97061202; PubMed=8905232;

Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horichi T.;

"A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";

DNA Res. 3:137-155(1996).

[4]

CHARACTERIZATION, AND SEQUENCE OF 496-510.

MEDLINE=96198179; PubMed=8626304;

Sturr M.G., Krulwich T.A., Hicks D.B.;

"Purification of a cytochrome bd terminal oxidase encoded by the *Escherichia coli* app locus from a delta cyo delta cyd strain complemented by genes from *Bacillus firmus* OF4.";

J. Bacteriol. 178:1742-1749(1996).

-1- FUNCTION: CYTOCHROME OXIDASE SUBUNIT.

-1- PATHWAY: Respiratory chain; terminal step.

-1- SUBUNIT: Heterodimer of subunits I and II (Probable).

-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (Probable).

-1- INDUCTION: IS INDUCED WHEN BACTERIAL CULTURES REACH STATIONARY PHASE; ITS SYNTHESIS IS TRIGGERED BY PHOSPHATE STARVATION OR A SHIFT FROM AEROBIC TO ANAEROBIC CONDITIONS.

-1- PTM: The N-terminus is blocked.

-1- SIMILARITY: STRONG, TO E.COLI CYD4.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; S63811; AAB20284.1; --

EMBL; AE000019; AAC74063.1; --

EMBL; D90735; BAA35743.1; --

PIR; S17958; S17958.

Ecogene; EGI1380; appC.

InterPro; IPR002585; Bac Ubp Cox.

Pfam; PF01654; Bac Ubp Cox; 1

Oxidoreductase; Electron transport; Transmembrane; Inner membrane; Heme; Complete proteome.

DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).

TRANSSEM 23 42 POTENTIAL.

DOMAIN 43 94 PERIPLASMIC (POTENTIAL).

TRANSSEM 95 114 POTENTIAL.

DOMAIN 115 129 CYTOPLASMIC (POTENTIAL).

TRANSSEM 130 149 POTENTIAL.

DOMAIN 150 187 PERIPLASMIC (POTENTIAL).

TRANSSEM 188 207 POTENTIAL.

DOMAIN 208 219 CYTOPLASMIC (POTENTIAL).

TRANSSEM 220 239 POTENTIAL.

DOMAIN 240 392 PERIPLASMIC (POTENTIAL).

TRANSSEM 393 412 POTENTIAL.

DOMAIN 413 470 CYTOPLASMIC (POTENTIAL).

TRANSSEM 471 490 POTENTIAL.

DOMAIN 514 514 PERIPLASMIC (POTENTIAL).

METAL 186 186 IRON (HEME AXIAL LIGAND).

SEQUENCE 514 AA; 57920 MW; 2D2FBD43429D960D CRC64; (BY SIMILARITY).

Query Match 29.6%; Score 50; DB 1; Length 514;

Best Local Similarity 50.0%; Pred. No. 30;

Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 16 WALALLMALVPPGW 29

DB 428 WSLPLFWTAIEAGW 441

RESULT 12

DPM3 HUMAN STANDARD; PRT; 92 AA.

AC Q9P2X0; Q9BXN4; Q9BXN5;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Dolichol-phosphate mannosyltransferase subunit 3 (Dolichol-phosphate mannosyltransferase subunit 3) (Dolichyl-phosphate beta-D-mannosyltransferase subunit 3) (Mannose-P-dolichol synthase subunit 3) (DPM synthase subunit 3) (DPM synthase complex subunit 3) (Prostin 1).

GN DPM3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_Taxid=9606;

[1]

RP SEQUENCE FROM N.A. (ISOFORM 1), SEQUENCE OF 1-13, AND CHARACTERIZATION.

RX MEDLINE=20296673; PubMed=10835346;

RA Maeda Y., Tanaka S., Hino J., Kangawa K., Kinoshita T.;

RT "Human dolichol-phosphate-mannose synthase consists of three subunits, DPM1, DPM2 and DPM3.";

RL EMBO J. 19:2475-2482(2000).

[2]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).

RX MEDLINE=21313795; PubMed=11420690;

RA Manos E.J., Kim M.L., Kassir J., Chang P.Y., Welle A., Jones D.A.;

RT "Dolichol-phosphate-mannose-3 (DPM3)/prostin-1 is a novel phospholipase C-gamma regulated gene negatively associated with prostate tumor invasion.";

RL Oncogene 20:2781-2790(2001).

CC -1- FUNCTION: Stabilizer subunit of the dolichol-phosphate-mannose synthase complex.

CC -1- SUBUNIT: Composed of three subunits; DPM1, DPM2 and DPM3.

CC Associated with DPM1 via its C-terminal domain and with DPM2 via its N-terminal portion.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=1; Synonyms=Short;

CC IsoId=Q9P2X0-1; Sequence=Displayed;

CC Name=2; Synonyms=Long;

CC IsoId=Q9P2X0-2; Sequence=VSP 001308;

CC -1- SIMILARITY: Belongs to the DPM3 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB028128; BAA96291.1; --

EMBL; AF312922; AAK28487.1; --

EMBL; AF312923; AAK28486.1; --

GeneW; HGNC:3007; DPM3.

MTM; 605951; --

DR Transmembrane; Endoplasmic reticulum; Alternative splicing.

KW TRANSSEM 8 28 POTENTIAL.

FT TRANSSEM 37 57 POTENTIAL.

FT VARSPLIC 1 1 M -> MLSVGLRLSLVRSFLLRGALLPSLAVTM (in

```

FT FT isoform 2).
FT FT /FTID=VSP_001308.
SQ SEQUENCE 92 AA; 10080 MW; C350B3696842A877 CRC64;
Query Match
Best Local Similarity 45.0%; Score 49.5; DB 1; Length 92;
Matches 9; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

2y 12 ASWCWALAL--WLAVVG 28
Db 5 AQMLMGALIGSTWALTTG 24

RESULT 13
PEXD CAEEL STANDARD; PRT; 330 AA.
AC Q19951.
2T 15-DEC-1998 (Rel. 37, Created)
2T 28-FEB-2003 (Rel. 41, Last sequence update)
2T 28-FEB-2003 (Rel. 41, Last annotation update)
2E Probable peroxisomal membrane protein PEX13 (Peroxin-13).
2N F32A5.6.
2S Caenorhabditis elegans.
2C Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
2C Rhabditidae; Pelodierinae; Caenorhabditis.
2X NCBI_TaxID=6239;
2N SEQUENCE FROM N.A.
2P STRAIN=Bristol NZ;
2L Pauley A.;
2L Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
2P REVISIONS.
2P Waterston R.;
2L Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
2C -1- FUNCTION: Component of the peroxisomal translocation machinery
2C with PEX14 and PEX17. Functions as a docking factor for the
2C predominantly cytoplasmic PEX1 receptor (PAX10/PEX5) (By
2C similarity).
2C -1- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal (By
2C similarity).
2C -1- SIMILARITY: Contains 1 SH3 domain.
2C -----
2C This SWISS-PROT entry is copyright. It is produced through a collaboration
2C between the Swiss Institute of Bioinformatics and the EMBL outstation -
2C the European Bioinformatics Institute. There are no restrictions on its
2C use by non-profit institutions as long as its content is in no way
2C modified and this statement is not removed. Usage by and for commercial
2C entities requires a license agreement (See http://www.isb-sib.ch/announce/
2C or send an email to license@isb-sib.ch).
2C -----
2C R EMBL; U02064; AAC4665.2; -.
2C R HSSP; P00519; 1ABU.
2C R WormBep; F32A5.6; CE28290.
2C R InterPro; IPR007223; Peroxin-13_N.
2C R Pfam; PF04088; Peroxin-13_N; 1.
2C R Pfam; PF00018; SH3; 1.
2C R PRINTS; PR00452; SH3DOMAIN.
2C R ProDom; PD000066; SH3; 1.
2C R SMART; SM00326; SH3; 1.
2C R PROSITE; PS50002; SH3; 1.
2C R Peroxisome; Transport; Protein transport; Transmembrane; SH3 domain.
2T DOMAIN 1 192 LIDENAL (POTENTIAL).
2T TRANSMEM 193 213 CYTOPLASMIC (POTENTIAL).
2T DOMAIN 214 330
2T DOMAIN 235 300 SH3.
2T SEQUENCE 330 AA; 35765 MW; 7DD647F07A04403 CRC64;

Query Match
Best Local Similarity 44.0%; Score 49.5; DB 1; Length 330;
Matches 11; Conservative 2; Mismatches 11; Indels 1; Gaps 1;

```

```

Qy 6 GARGASWCWALALMLAVVG-W 29
Db 187 GATRPASVNPALFWVVAIGPW 211

RESULT 14
INT SYNY3 STANDARD; PRT; 519 AA.
AC INT SYNY3
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apolipoprotein N-acyltransferase (EC 2.3.1.-) (ALP N-acyltransferase).
GN LNT OR SLR0819.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
2N SEQUENCE FROM N.A.
2P MEDLINE=97061201; PubMed=8905231;
2C Kaneko T., Sato S., Kocani H., Tanaka A., Asamizu E., Nakamura Y.,
2C Miyajima N., Hikosawa M., Sugitara M., Sasamoto S., Kimura T.,
2C Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Natsu K., Okamura S.,
2C Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
2C Tabata S.;
2C "Sequence analysis of the genome of the unicellular cyanobacterium
2C Synechocystis sp. strain PCC6803. II. Sequence determination of the
2C entire genome and assignment of potential protein-coding regions.";
2C DNA Res. 3:109-136(1996).
2C -1- FUNCTION: Transfers the fatty acyl group on membrane lipoproteins
2C (By similarity).
2C -1- PATHWAY: Lipoproteins biosynthesis.
2C -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
2C -1- SIMILARITY: Belongs to the apolipoprotein N-acyltransferase
2C family.
2C -1- SIMILARITY: Contains 1 CN hydrolase domain.
2C -----
2C This SWISS-PROT entry is copyright. It is produced through a collaboration
2C between the Swiss Institute of Bioinformatics and the EMBL outstation -
2C the European Bioinformatics Institute. There are no restrictions on its
2C use by non-profit institutions as long as its content is in no way
2C modified and this statement is not removed. Usage by and for commercial
2C entities requires a license agreement (See http://www.isb-sib.ch/announce/
2C or send an email to license@isb-sib.ch).
2C -----
2C DR EMBL; D90911; BAA1813.1; -.
2C DR PIR; S75570; S75570.
2C DR InterPro; IPR004563; Lnt.
2C DR InterPro; IPR003010; Ntlse/CNhydase.
2C DR Pfam; PR00795; CN hydrolase; 1.
2C DR TIGRfam; TIGR00546; lnt; 1.
2C DR PROSITE; PS50263; CN HYDROLASE; 1.
2C R Transferase; Acyltransferase; Transmembrane; Complete proteome.
2T TRANSFER 6 26
2T TRANSMEM 47 67 POTENTIAL.
2T TRANSMEM 83 103 POTENTIAL.
2T TRANSMEM 126 146 POTENTIAL.
2T TRANSMEM 174 194 POTENTIAL.
2T TRANSMEM 206 226 POTENTIAL.
2T TRANSMEM 277 297 POTENTIAL.
2T TRANSMEM 496 516 POTENTIAL.
2T DOMAIN 239 519 CN HYDROLASE.
2T SEQUENCE 519 AA; 58130 MW; A28C75F260E8BBA3 CRC64;

Query Match
Best Local Similarity 38.5%; Score 49.5; DB 1; Length 519;
Matches 10; Conservative 3; Mismatches 2; Indels 11; Gaps 2;

```



```

RESULT 15
LN8 HUMAN
LN8 HUMAN STANDARD; PRT; 286 AA.
C Q9UB8; Q96195;
C 28-FEB-2003 (Rel. 41, Created)
C 28-FEB-2003 (Rel. 41, Last sequence update)
C 10-OCT-2003 (Rel. 42, Last annotation update)
LN8 CLN8 protein.
LN8 CLN8.
LN8 Homo sapiens (Human).
LN8 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
LN8 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
LN8 NCBI_TaxID=9606;
LN8 [1]
LN8 SEQUENCE FROM N.A., VARIANT EPMR GLY-24, AND VARIANT ALA-155.
LN8 MEDLINE=99438402; PubMed=10508524;
LN8 Ranta S., Zhang Y., Ross B., Lonka L., Takunen E., Messer A.,
LN8 Sharp J., Wheeler R., Kusumi K., Mole S., Liu W., Soares M.B.,
LN8 Bonaldo M.F., Hirvasniemi A., de la Chapelle A., Gilliam T.C.,
LN8 Lehesjoki A.-E.;
LN8 "The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice
LN8 are associated with mutations in CLN8.";
LN8 Nat. Genet. 23:233-236(1999).
LN8 [2]
LN8 SEQUENCE FROM N.A.
LN8 TISSUE=Uterus;
LN8 MEDLINE=22388257; PubMed=12477932;
LN8 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
LN8 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
LN8 Altschul S.P., Zeeberg B., Bluetow K.H., Schaefer C.F., Bhat N.K.,
LN8 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
LN8 Datchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
LN8 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
LN8 Brownstein M.J., Ustin T.B., Toshynski A., Canninci P., Prange C.,
LN8 Raba S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
LN8 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
LN8 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
LN8 Vallilo D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
LN8 Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
LN8 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
LN8 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
LN8 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
LN8 Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
LN8 Scherch A., Schein J.E., Jones S.J.M., Maira M.A.;
LN8 "Generation and initial analysis of more than 15,000 full-length
LN8 human and mouse cDNA sequences.";
LN8 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
LN8 [3]
LN8 SUBCELLULAR LOCATION, AND MUTAGENESIS OF 283-LYS-LYS-284.
LN8 MEDLINE=20320699; PubMed=10861296;
LN8 Lonka L., Kyttaelae A., Ranta S., Jalanko A., Lehesjoki A.-E.;
LN8 "The neuronal ceroid lipofuscinoses CLN8 membrane protein is a
LN8 resident of the endoplasmic reticulum.";
LN8 Hum. Mol. Genet. 9:1691-1697(2000).
LN8 -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
LN8 reticulum and ER-Golgi intermediate compartment (ERGIC).
LN8 -1- PTM: Does not seem to be N-glycosylated.
LN8 -1- DISEASE: Defects in CLN8 are a cause of progressive epilepsy with
LN8 mental retardation (EPMR) [MIM:601493]; also known as neuronal
LN8 ceroid lipofuscinoses type 8 and Northern epilepsy. EPMR is an
LN8 autosomal recessive disorder characterized by normal early
LN8 development, onset of generalized seizures between 5 and 10 years,
LN8 and subsequent progressive mental retardation. Biochemically, the
LN8 disease is characterized by the intracellular accumulation of
LN8 hydrophobic material, mainly ATP synthase subunit C.
LN8 -1- SIMILARITY: Contains 1 TIG (TRAM/LAG1/CLN8) domain.
LN8 -1- DATABASE: NAME=NCL CLN8;
LN8 NOTE=Neural Ceroid Lipofuscinoses mutation db;
LN8 WWW="http://www.ucl.ac.uk/ncl/CLN8.html".

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF123757; AAF13115.1; -
DR EMBL: AF123758; AAF13116.1; -
DR EMBL: AF123759; AAF13117.1; -
DR EMBL: AF123760; AAF13118.1; -
DR EMBL: AF123761; AAF13119.1; -
DR EMBL: BC007725; AA07725.1; -
DR Genew: HGNC:2079; CLN8.
DR MIM: 607837; -
DR MIM: 600143; -
DR GO: GO:0005783; C:endoplasmic reticulum; TAS.
DR GO: GO:0005793; C:ER-Golgi intermediate compartment; TAS.
DR GO: GO:0016021; C:integral to membrane; TAS.
DR GO: GO:0007399; P:neurogenesis; TAS.
DR InterPro: IPR006634; TIG.
DR SMART: SM00724; TIG; 1.
DR PROSITE: PS05922; TIG; 1.
KW Transmembrane; Endoplasmic reticulum; Neuronal ceroid lipofuscinoses;
KW Disease mutation; Polymorphism; Epilepsy.
FT TRANSMEM 21 41
FT TRANSMEM 62 84
FT TRANSMEM 103 123
FT TRANSMEM 131 151
FT TRANSMEM 226 246
FT DOMAIN 62 262
FT SITE 283 286
FT VARIANT 24 24
FT VARIANT 155 155
FT VARIANT 283 284
FT MUTAGEN 225 225
FT CONFLICT 225 225
SQ
SQ SEQUENCE 286 AA; 32842 MW; 0BAEBCA516D1DC CRC64;
Query Match 29.0%; Score 49; DB 1; Length 286;
Best Local Similarity 58.3%; Pred. No. 24;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
DB 93 AGRARSGQNCW 104

```

Search completed: June 8, 2004, 14:04:38
Job time : 1.25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

on: June 8, 2004, 13:59:08 ; Search time 5.54167 seconds

(without alignments)
1708.068 Million cell updates/sec

File: US-10-010-050a-2_COPY_1_30

Sequence: 1 MRRGGAARGRASWCALALMLAVVPGWS 30

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Marched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mmc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_protent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_ivirus:*
16: SP_bacteriaph:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	39.1	606	17	Q9HMT4
2	65.5	38.6	205	16	Q8PDE8
3	63.5	37.6	201	16	Q8PDE5
4	62	36.7	588	16	Q82G94
5	59	34.9	198	16	Q8PLY4
6	59	34.9	198	16	Q8PA94
7	59	34.9	304	16	Q7WFK8
8	59	34.9	304	16	Q7W451
9	58	34.3	412	16	Q92VB2
10	58	34.3	146	11	Q8CEK6
11	57.5	34.0	343	16	Q8G419
12	57	33.7	226	16	Q8DG07
13	57	33.7	366	16	Q84X72
14	56	33.1	214	16	Q69464
15	56	33.1	485	16	Q88SV3
16	55.5	32.8	371	4	Q96H96

17	55.5	32.8	412	10	Q7XSY0	Q7XSY0 oryza sativ
18	55.5	32.8	695	16	Q8XW44	Q8XW44 ralsionia s
19	55	32.5	147	10	Q7X917	Q7X917 oryza sativ
20	55	32.5	232	16	Q9PEH4	Q9PEH4 xyella fas
21	55	32.5	232	16	Q87EH4	Q87EH4 xyella fas
22	55	32.5	707	10	Q9ZRA6	Q9ZRA6 chlamydomon
23	55	32.5	836	16	Q89HK3	Q89HK3 brachyrihob
24	55	32.5	3019	12	Q92529	Q92529 hepatitis c
25	54.5	32.2	186	10	Q8LNI9	Q8LNI9 oryza sativ
26	54.5	32.2	435	10	Q9XE69	Q9XE69 sorghum bic
27	54	32.0	54	8	Q9WIA5	Q9WIA5 typhlonec
28	54	32.0	214	10	Q9XE06	Q9XE06 sorghum bic
29	54	32.0	225	10	Q84ZK3	Q84ZK3 oryza sativ
30	54	32.0	246	16	Q87AE0	Q87AE0 xyella fas
31	54	32.0	666	2	Q8L311	Q8L311 vitreoscill
32	53.5	31.7	126	2	Q8WV97	Q8WV97 alcaligenes
33	53.5	31.7	284	16	Q8WV99	Q8WV99 xyella fas
34	53.5	31.7	284	16	Q87EH8	Q87EH8 xyella fas
35	53.5	31.7	749	16	Q825X8	Q825X8 nitrosomona
36	53	31.4	83	9	Q7Y4J5	Q7Y4J5 streptococc
37	53	31.4	98	17	Q8ZSP9	Q8ZSP9 pyrobaculum
38	53	31.4	160	16	Q82HM2	Q82HM2 streptomyce
39	53	31.4	224	2	Q8EY24	Q8EY24 xanthomonas
40	53	31.4	428	16	Q8UD14	Q8UD14 agrobacteri
41	53	31.4	455	16	Q88D16	Q88D16 pseudomonas
42	53	31.4	504	16	Q7WVW8	Q7WVW8 bordetella
43	52.5	31.1	504	16	Q7VXU4	Q7VXU4 bordetella
44	52.5	31.1	220	16	Q8YBF6	Q8YBF6 brucella me
45	52.5	31.1	325	16	Q8FWY7	Q8FWY7 brucella su

ALIGNMENTS

RESULT 1

Q9HMT4 PRELIMINARY; PRT; 606 AA.

AC Q9HMT4 01-MAR-2001 (TREMblrel. 16, Created)

DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)

DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)

DE Sulfate transport system permease protein.

GN CYST1 OR VNG2396G.

OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).

OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Halobacterium.

OX NCBI_TaxID=64091;

RN [1]

RP SEQUENCE FROM N. A.

RX MEDLINE=20504483; PubMed=11016950;

RA Ng W.V., Kennedy S.P., Mahatras G.G., Bergquist B., Pan M., Shukla H.D., Lasky S.R., Baliga N.S., Thorsen V., Shrotra J., Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A., Leithausen B., Keller K., Cruz R., Danson M.T., Hough D.W., Ikenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H., Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D., Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Daasarna S., "Genome sequence of Halobacterium species NRC-1"; Proc. Natl. Acad. Sci. U.S.A. 97:12175-12181(2000).

CC -FUNCTION: PROBABLY PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE (BY SIMILARITY).

CC -SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

CC -SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEMS.

EMBL: AF005121; AAC20487.1; -.

DR PIR; C84390; C84390.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0003700; F:transcription factor activity; IEA.

DR GO; GO:0005215; F:transporter activity; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006810; P:transport; IEA.

```

JR InterPro: IPR000515, BPD_transp.
JR InterPro: IPR000847, HTH_LysR.
JR Pfam: PF00528, BPD_transp; 2.
JR PROSITE: PS00402, BPD_TRANS INN_MEMBER; 1.
JR PROSITE: PS00044, HTH_LYSR_FAMILY; 1.
JR Transmembrane: Transport; Complete proteome.
NO SEQUENCE 606 AA; 6306 MW; 26489A20298A28C CRC64;

Query Match      39.1%; Score 66; DB 17; Length 606;
Best Local Similarity 53.8%; Pred. No. 2.1;
Matches 14; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

b      3  RGAGAGRGRASWCALMLLWAVPG 28
      |||:|||||:|||||:|||||:
      508  RALGASGRRAIMVDELPLVWGVAG 533

ESULT 2
C QBPDP8      PRELIMINARY;      PRT;      205 AA.
D QBPDP8;
T 01-OCT-2002 (TREMBlrel. 22, Created)
T 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
T 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
E Hypothetical protein XCC0380.
N XCC0380.
C Xanthomonas campestris (pv. campestris).
C Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
C Xanthomonadaceae; Xanthomonas.
N NCBI_TaxID=340;
[1]
P SEQUENCE FROM N.A.
C STRAIN=ATCC 33913 / NCPPB 528;
X MEDLINE=22022145; PubMed=12024217;
da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
Camarote G., Camavan F., Cardozo J., Chamberg F., Chapina L.P.,
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
Faria J.B., Ferreira A.U.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
Setubal J.C., Kitajima J.P.
"Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
Nature 417:459-463(2002).
EMBL: AE012134; AAM33699.1; -.
InterPro: IPR003848; DUF218.
Pfam: PR02698; DUF218; 1.
Hypothetical protein; Complete proteome.
NO SEQUENCE 205 AA; 23048 MW; CCC1D7C8BAF9A482 CRC64;

Query Match      38.8%; Score 65.5; DB 16; Length 205;
Best Local Similarity 42.4%; Pred. No. 0.92;
Matches 14; Conservative 3; Mismatches 11; Indels 5; Gaps 2;

/      2  RRGAGAGRGRASWCMA-----LALLMLAVVPGM 29
      |||:|||||:|||||:|||||:
      3  RRVSRPRSGLGLMGWGRMLCMALIMLVAVGM 35

ESULT 3
C QBPDP8      PRELIMINARY;      PRT;      201 AA.
D QBPDP8;
T 01-OCT-2002 (TREMBlrel. 22, Created)
T 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

```

```

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein XAC0380.
GN XAC0380.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
N NCBI_TaxID=92829;
[1]
P SEQUENCE FROM N.A.
C STRAIN=306 / ATCC 13902 / XV 101;
X MEDLINE=22022145; PubMed=12024217;
da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
Camarote G., Camavan F., Cardozo J., Chamberg F., Chapina L.P.,
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
Faria J.B., Ferreira A.U.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
Setubal J.C., Kitajima J.P.
"Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
Nature 417:459-463(2002).
EMBL: AE011664; AAM35272.1; -.
InterPro: IPR003848; DUF218.
DR Pfam: PR02698; DUF218; 1.
KW Hypothetical protein; Complete proteome.
NO SEQUENCE 201 AA; 22724 MW; 62D3709290F3D458 CRC64;

Query Match      37.6%; Score 63.5; DB 16; Length 201;
Best Local Similarity 44.4%; Pred. No. 1.6;
Matches 12; Conservative 2; Mismatches 8; Indels 5; Gaps 1;

QY      8  AGRASWCMA-----LALLMLAVVPGM 29
      |||:|||||:|||||:|||||:
DB      5  SRGLMGWGRMLCMALIMLVAVGM 31

RESULT 4
C QBPDP8      PRELIMINARY;      PRT;      588 AA.
D QBPDP8;
T 01-JUN-2003 (TREMBlrel. 24, Created)
T 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
T 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
E Hypothetical protein.
N SAV4004.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
N NCBI_TaxID=33903;
[1]
P SEQUENCE FROM N.A.
C STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
X MEDLINE=21477403; PubMed=11572948;
Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
Shinohe M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
Kikuchi H., Shiba T., Sakaki Y., Hattori M.,
"Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites."
Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
[2]
P SEQUENCE FROM N.A.
C STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
X MEDLINE=22608306; PubMed=12692562;
Ikeda H., Ishikawa J., Hanamoto A., Shinohe M., Kikuchi H., Shiba T.,

```

A Sakaki Y., Hattori M., Omura S.;
 I "Complete genome sequence and comparative analysis of the industrial
 L microorganism *Streptomyces avermitilis*."; Nat. Biotechnol. 21:526-531(2003).
 R EMBL; AP005037; BAC71716.1; -;
 M Hypothetical protein; Complete proteome.
 N SEQUENCE 588 AA; 63428 MW; D8DC5F23BE49A321 CRC64;

Query Match 36.7%; Score 62; DB 16; Length 588;
 Best Local Similarity 28.6%; Pred. No. 6.7;
 Matches 14; Conservative 5; Mismatches 8; Indels 22; Gaps 1;

3 RGAGAAAR-----GRASWCWALTLMLAVVPGW 29
 167 RGAGSRTLVATKNTLIGVYGRYGDKSMWAGTINLAPARW 215

RESULT 5
 PRELIMINARY; PRT; 198 AA.

08PLY4
 01-OCT-2002 (TREMBlrel. 22, Created)
 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 TonB-like protein.

Xanthomonas axonopodis (pv. citri).
 Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 Xanthomonadaceae; Xanthomonas.
 NCBI_TaxID=92829;

SEQUENCE FROM N.A.

STRAIN=306 / ATCC 13902 / XV 101;
 MEDLINE=22022145; PubMed=12024217;

da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 Camarotte G., Cannavan F., Cardoso J., Chamberg F., Clapina L.P.,
 Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 Setubal J.C., Kitajima J.P.;
 "Comparison of the genomes of two Xanthomonas pathogens with differing
 host specificities."; Nature 417:459-463(2002).
 EMBL; AF011797; AAM36519.1; -;
 InterPro; IPR006260; TonB_C.
 TIGRFAMs; TIGR01352; TonB_Cterm; 1.
 Complete proteome.
 SEQUENCE 198 AA; 21748 MW; FB9DF173EDC52961 CRC64;

Query Match 34.9%; Score 59; DB 16; Length 198;
 Best Local Similarity 43.5%; Pred. No. 6.1;
 Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

7 AARGASWCWALTLMLAVVPGW 29
 36 ASDGASPRMWTIVLVVPSW 58

PRELIMINARY; PRT; 198 AA.

01-OCT-2002 (TREMBlrel. 22, Created)
 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE TonB-like protein.
 GN XCC1592.
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 NCBI_TaxID=340;

SEQUENCE FROM N.A.
 STRAIN=ATCC 33913 / NCPPB 528;
 MEDLINE=22022145; PubMed=12024217;

da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 Camarotte G., Cannavan F., Cardoso J., Chamberg F., Clapina L.P.,
 Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 Setubal J.C., Kitajima J.P.;
 "Comparison of the genomes of two Xanthomonas pathogens with differing
 host specificities."; Nature 417:459-463(2002).
 DR EMBL; AE012260; AAM40887.1; -;
 DR InterPro; IPR006260; TonB_C.
 TIGRFAMs; TIGR01352; TonB_Cterm; 1.
 Complete proteome.
 SEQUENCE 198 AA; 21744 MW; 78A3626C6142C60C CRC64;

Query Match 34.9%; Score 59; DB 16; Length 198;
 Best Local Similarity 50.0%; Pred. No. 6.1;
 Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

10 GRASWCWALTLMLAVVPGW 29
 41 GASPMWATVLAIVVPSW 60

PRELIMINARY; PRT; 304 AA.

01-OCT-2003 (TREMBlrel. 25, Created)
 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 Putative membrane protein.
 BN4263.

Bordetella bronchiseptica (Alcaligenes bronchiseptica).
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 NCBI_TaxID=518;

SEQUENCE FROM N.A.

STRAIN=RB50 / ATCC BAA-588;
 MEDLINE=22827954; PubMed=12910271;

Parkhill U., Sebatina M., Preston A., Murphy L.D., Thomson N.,
 Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagsels K.,
 Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 Sharp S., Simmonds M., Skelton J., Squares R., Squares K.,
 Unwin L., Whitehead S., Barrall B.G., Maskell D.J.;
 "Comparative analysis of the genome sequences of *Bordetella pertussis*,
Bordetella parapertussis and *Bordetella bronchiseptica*.";

```

L1 Nat. Genet. 35:32-40(2003).
L2 EMBL; BX640450; CAE34627.1; -.
L3 Complete proteome.
L4 SEQUENCE 304 AA; 31640 MW; 7AF46DA2PB95EFC0 CRC64;
L5
L6 Query Match 34.9%; Score 59; DB 16; Length 304;
L7 Best Local Similarity 39.5%; Pred. No. 9;
L8 Matches 15; Conservative 2; Mismatches 9; Indels 12; Gaps 2;
L9
L10
L11
L12
L13
L14
L15
L16
L17
L18
L19
L20
L21
L22
L23
L24
L25
L26
L27
L28
L29
L30
L31
L32
L33
L34
L35
L36
L37
L38
L39
L40
L41
L42
L43
L44
L45
L46
L47
L48
L49
L50
L51
L52
L53
L54
L55
L56
L57
L58
L59
L60
L61
L62
L63
L64
L65
L66
L67
L68
L69
L70
L71
L72
L73
L74
L75
L76
L77
L78
L79
L80
L81
L82
L83
L84
L85
L86
L87
L88
L89
L90
L91
L92
L93
L94
L95
L96
L97
L98
L99

```

```

RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puhler A.;
RT "The complete sequence of the 1,683-kb pSYMB megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AF603644; CAC49196.1; -.
DR PIR; D95941; D95941.
DR GO; GO:0046821; Cytochromeosomal DNA; IEA.
DR GO; GO:0005489; Electron transporter activity; IEA.
DR GO; GO:0006118; Electron transport; IEA.
DR InterPro; IPR000345; Cyt c heme BS.
DR PROSITE; PS00190; CYTOCHROME_C_1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 412 AA; 46126 MW; 8088B8B0C66F838 CRC64;
L1
L2
L3
L4
L5
L6
L7
L8
L9
L10
L11
L12
L13
L14
L15
L16
L17
L18
L19
L20
L21
L22
L23
L24
L25
L26
L27
L28
L29
L30
L31
L32
L33
L34
L35
L36
L37
L38
L39
L40
L41
L42
L43
L44
L45
L46
L47
L48
L49
L50
L51
L52
L53
L54
L55
L56
L57
L58
L59
L60
L61
L62
L63
L64
L65
L66
L67
L68
L69
L70
L71
L72
L73
L74
L75
L76
L77
L78
L79
L80
L81
L82
L83
L84
L85
L86
L87
L88
L89
L90
L91
L92
L93
L94
L95
L96
L97
L98
L99

```

Bifidobacterium longum.
 Bacteria; Actinobacteriia; Actinobacteridae; Bifidobacteriales;
 Bifidobacteriaceae; Bifidobacterium.
 NCBI_TaxID=216816;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=NC 2705;
 MEDLINE=22294977; PubMed=12381787;
 Schell M.A., Karmaliantzou M., Snel B., Vilanova D., Berger B.,
 Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
 Pirmore R.D., Arigoni F.;
 "The genome sequence of Bifidobacterium longum reflects its adaptation
 to the human gastrointestinal tract."
 Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
 EMBL; AB014793; AAN25367.1; -
 Hypothetical protein; Complete proteome.
 SEQUENCE 343 AA; 37507 MW; 8AB35FF32D2228059 CRC64;

Query Match 34.0%; Score 57.5; DB 16; Length 343;
 Best Local Similarity 54.2%; Pred. No. 16;
 Matches 13; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

9 RGR-----ASNCMALALLMLAVP 27
 266 RGKCSYVAASWVMVALLMLAAYP 289

35007 12
 35007 PRELIMINARY; PRT; 226 AA.
 08DQ07;
 01-MAR-2003 (TRENBLrel. 23, Created)
 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 TLR2258 protein.
 TLR2258.
 Synecchococcus elongatus (Thermosynechococcus elongatus).
 Bacteria; Cyanobacteriia; Chroococcales; Synecchococcus.
 NCBI_TaxID=32046;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=BP-1;
 MEDLINE=22225144; PubMed=12240834;
 Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 Wakamatsu Y., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,
 Kiyokawa C., Kohara M., Matsunoto M., Matsuno A., Nakazaki N.,
 Shimpo S., Sugiimoto M., Takeuchi C., Yamada M., Tabata S.;
 "Complete genome structure of the thermophilic cyanobacterium
 Thermosynechococcus elongatus BP-1."
 DNA Res. 9:123-130(2002).
 EMBL; AP005376; BAC09810.1; -
 Complete proteome.
 SEQUENCE 226 AA; 25237 MW; 3A176B4B3792DC2 CRC64;

Query Match 33.7%; Score 57; DB 16; Length 226;
 Best Local Similarity 41.4%; Pred. No. 12;
 Matches 12; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

2 RGAAGARGASWCMALALLMLAVP 30
 3 RGGWTECGWVWGCLALLMLPLSSAMS 31

35007 13
 34X72 PRELIMINARY; PRT; 366 AA.
 084X72;
 01-JUN-2003 (TRENBLrel. 24, Created)
 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 CR66 protein.
 Chlamydomonas reinhardtii.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

OC Chlamydomonadaceae; Chlamydomonas.
 OC NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Li J.B., Lin S., Jia H., Wu H., Roe B.A., Kulp D., Stormo G.D.,
 RA Dutcher S.K.;
 RT "Finished genomic sequence in Chlamydomonas";
 RL Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AY207499; AA032622.1; -
 SEQUENCE 366 AA; 38876 MW; 99F5238B9E625BB CRC64;

Query Match 33.7%; Score 57; DB 10; Length 366;
 Best Local Similarity 45.5%; Pred. No. 19;
 Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

4 GAGAARGASWCMALALLMLAV 25
 315 GGGGQPGRRARCMAGFLWVEV 336

RESULT 14
 069464 PRELIMINARY; PRT; 214 AA.
 ID 069464;
 AC 069464;
 DT 01-AUG-1998 (TRENBLrel. 07, Created)
 DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Putative integral membrane protein (Conserved integral membrane
 DE protein).
 GN MLI666 OR MLCB1243.07.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OK NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eigemeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Whittall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
 RA Holtroyd S., Hornsby T., Jagele K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrall B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 DR EMBL; AL023635; CAI19188.1; -
 DR EMBL; AL583923; CAC30619.1; -
 DR PIR; T44701; T44701.
 DR Lepidoma; MLI666; -
 KW Complete proteome.
 SQ SEQUENCE 214 AA; 23402 MW; 5B48414709828FA0 CRC64;

Query Match 33.1%; Score 56; DB 16; Length 214;
 Best Local Similarity 41.7%; Pred. No. 16;
 Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

3 RGAAGARGASWCMALALLMLAV 26
 175 RTSAIARGIYOMKRSITIMFTIV 198

RESULT 15
 088SV3 PRELIMINARY; PRT; 485 AA.
 ID 088SV3;
 AC 088SV3;
 DT 01-JUN-2003 (TRENBLrel. 24, Created)
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Cardiolipin synthetase 2 (EC 2.7.8.-).

```

2N CLS OR LP 3273.
2S Lactobacillus plantarum.
2C Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
2X Lactobacillus.
2Y NCBI_TaxID=1590;
2Z [1]
2P SEQUENCE FROM N.A.
2C STRAIN=NCIMB 8826 / WCFS1;
2X MEDLINE=22480296; PubMed=12566566;
2A Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
2A Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
2A Eilers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
2A Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
2A De Vos W.M., Siezen R.J.;
2T "Complete genome sequence of Lactobacillus plantarum WCFS1.";
2L Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
2R EMBL; AL935261; CAD65406.1; -.
2R GO; GO:0016740; F:transferase activity; IEA.
2R GO; GO:0008152; P:metabolism; IEA.
2R InterPro: IPR01736; PLD.
2R Pfam; PF00614; PLDc; 2.
2R PROSITE; PS50035; PLD; 2.
2M Transferase; Complete Proteome.
3Q SEQUENCE 485 AA; 55778 MW; C725C28841DF8D31 CRC64;

```

Query Match 33.1%; Score 56; DB 16; Length 485;
 Best Local Similarity 46.7%; Pred. No. 33;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

2Y 12 ASWCALALMLAV 26
2b 36 ATWMLLVIMLPPI 50

```

Search completed: June 8, 2004, 14:07:06
 Job time : 5.54167 secs

reagents to detect zsi946 or cells expressing it, e.g. for assessing thyroid function to produce anti-idiotypic antibodies, for affinity purification of zsi946, to screen expression libraries, to neutralise zsi946 activity, and to deliver toxins, radioisotopes etc. for therapeutic or diagnostic purposes. Agonists of the product can be used to promote growth, differentiation and proliferation of specific cell types, e.g. for treating (extra)thyroid diseases or as additive to cell cultures

Sequence 346 AA:

Query Match 100.0%; Score 1758; DB 2; Length 346;
 Best Local Similarity 100.0%; Pred. No. 3.6e-169;
 Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 RVSGIPSRHRMPVYKRFDPKPPYCOAKYTPCGTSPPIVWBGDDIEVFRLOAPVW 60
 b RVSGIPSRHRMPVYKRFDPKPPYCOAKYTPCGTSPPIVWBGDDIEVFRLOAPVW 90
 Y 61 EFKYGDLLGHLKIMHDAIGFRSTLTGKNYTWEMWELFOLGNCCTPHLRPMDAPFWCNOG 120
 b 91 EFKYGDLLGHLKIMHDAIGFRSTLTGKNYTWEMWELFOLGNCCTPHLRPMDAPFWCNOG 150
 Y 121 AACFEFGIDDVHMKENTLVQVATISGNMNMOMAKVQKQNETGTYETWNVKASPEKGA 180
 b 151 AACFEFGIDDVHMKENTLVQVATISGNMNMOMAKVQKQNETGTYETWNVKASPEKGA 210
 Y 181 ETWFDSDYSCSKFVLRTEFNKLAEGAFKFNIEYTRIFLYSGEPTLYAGNETSVFGPTGNK 240
 b 211 ETWFDSDYSCSKFVLRTEFNKLAEGAFKFNIEYTRIFLYSGEPTLYAGNETSVFGPTGNK 270
 Y 241 TLGLAIKRFYYPKPHLPTEKEFLLSLLOIFDAVIVHKQFYLFFNFEYFWELPMKPPFIKIT 300
 b 271 TLGLAIKRFYYPKPHLPTEKEFLLSLLOIFDAVIVHKQFYLFFNFEYFWELPMKPPFIKIT 330
 Y 301 YEEIPLPIRNKTLISGL 316
 b 331 YEEIPLPIRNKTLISGL 346

RESULT 2
 AA41323
 D AA41323 standard; protein; 347 AA.
 C AA41323;
 X
 Y 02-DEC-1999 (first entry)
 E Human secreted protein encoded by gene 16 clone HMZAD77.
 K Human; secreted protein; fusion protein; gene therapy; protein therapy;
 M diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 M developmental abnormality; foetal deficiency; blood; allergy; renal;
 M immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 M inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 M cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 M osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 M endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 K Homo sapiens.
 K MO947540-A1.
 X
 Y 23-SEP-1999.
 K
 Y 18-MAR-1999; 99WO-US005804.
 K
 Y 19-MAR-1998; 98US-0078563P.
 X 19-MAR-1998; 98US-0078566P.
 X 19-MAR-1998; 98US-0078573P.
 X 19-MAR-1998; 98US-0078574P.
 X 19-MAR-1998; 98US-0078576P.
 X 19-MAR-1998; 98US-0078577P.

PR 19-MAR-1998; 98US-0078578P.
 PR 19-MAR-1998; 98US-0078579P.
 PR 19-MAR-1998; 98US-0078581P.
 PR 01-APR-1998; 98US-0080312P.
 PR 01-APR-1998; 98US-0080313P.
 PR 01-APR-1998; 98US-0080314P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PI Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR,
 PI Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;
 PI Shi Y, Moore PA;
 XX
 DR WPI: 1999-562050/47.
 XX N-FSDB; AA224826.
 PT
 PT New isolated human genes, useful for diagnosis and treatment of e.g.
 PT cancers, neurological disorders, immune diseases, inflammation or blood
 PT disorders.
 PS
 XX Claim 11; Page 367-368; 484pp; English.
 CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AA224802) for increasing the stability of the fused protein
 CC as compared to the human protein only. The invention relates to 95 novel
 CC genes and their fragments (nucleic acid sequences: AA224811-224907; amino
 CC acid sequences AA41308-41404) which are useful for preventing, treating
 CC or ameliorating medical conditions e.g. by protein or gene therapy. Also,
 CC pathological conditions can be diagnosed by determining the amount of the
 CC new polypeptides in a sample or by determining the presence of mutations
 CC in the new polynucleotides. Specific uses are described for each of the
 CC 95 polynucleotides, based on which tissues they are most highly expressed
 CC in (see AA224811 for described uses)
 CC
 XX
 XX SQ Sequence 347 AA;

Query Match 99.4%; Score 1747; DB 2; Length 347;
 Best Local Similarity 99.4%; Pred. No. 4.7e-168;
 Matches 314; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RVSGIPSRHRMPVYKRFDPKPPYCOAKYTPCGTSPPIVWBGDDIEVFRLOAPVW 60
 Db 31 RVSGIPSRHRMPVYKRFDPKPPYCOAKYTPCGTSPPIVWBGDDIEVFRLOAPVW 90
 QY 61 EFKYGDLLGHLKIMHDAIGFRSTLTGKNYTWEMWELFOLGNCCTPHLRPMDAPFWCNOG 120
 Db 91 EFKYGDLLGHLKIMHDAIGFRSTLTGKNYTWEMWELFOLGNCCTPHLRPMDAPFWCNOG 150
 QY 121 AACFEFGIDDVHMKENTLVQVATISGNMNMOMAKVQKQNETGTYETWNVKASPEKGA 180
 Db 151 AACFEFGIDDVHMKENTLVQVATISGNMNMOMAKVQKQNETGTYETWNVKASPEKGA 210
 QY 181 ETWFDSDYSCSKFVLRTEFNKLAEGAFKFNIEYTRIFLYSGEPTLYAGNETSVFGPTGNK 240
 Db 211 ETWFDSDYSCSKFVLRTEFNKLAEGAFKFNIEYTRIFLYSGEPTLYAGNETSVFGPTGNK 270
 QY 241 TLGLAIKRFYYPKPHLPTEKEFLLSLLOIFDAVIVHKQFYLFFNFEYFWELPMKPPFIKIT 300
 Db 271 TLGLAIKRFYYPKPHLPTEKEFLLSLLOIFDAVIVHKQFYLFFNFEYFWELPMKPPFIKIT 330
 QY 301 YEEIPLPIRNKTLISGL 316
 Db 331 YEEIPLPIRNKTLISGL 346

RESULT 3
 AAM93870
 ID AAM93870 standard; protein; 358 AA.
 XX
 AC AAM93870;
 XX

06-NOV-2001 (first entry)
 Human polypeptide, SEQ ID NO: 3978.
 Human; full length cDNA; cDNA synthesis; oligo-capping.
 Homo sapiens.
 EP130094-A2.
 05-SEP-2001.
 07-JUL-2000; 2000BP-00114089.
 08-JUL-1999; 99JP-00194486.
 11-JAN-2000; 2000JP-00118774.
 02-MAY-2000; 2000JP-00183765.
 (HELI-) HELIX RES INST.
 Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 WPI; 2001-524255/58.
 N-PSDB; AAK94829.

830 Primers useful for synthesizing full length cDNA clones and their use in genetic manipulation.

Claim 8; SEQ ID NO 3978; 1380bp + Sequence Listing; English.

The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a polypeptide encoded by a full length human cDNA of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO

Sequence 358 AA;

Query March 99.2%; Score 1744; DB 4; Length 358;

Best local Similarity 99.4%; Pred. No. 9.9e-168;

Matches 314; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

1  RVSGIPSRHMPVYKRRFDPKDPYCOAKYTCPTGSPPIPVMEGDDIEVFLQAPVW 60
43  RVSGIPSRHMPVYKRRFDPKDPYCOAKYTCPTGSPPIPVMEGDDIEVFLQAPVW 102
61  EFKYGDLLGHLKIHDAIGFRSTLTGKNYTWMEYFOLGNCETPHURPEMDAPFPCNOG 120
103  EFKYGDLLGHLKIHDAIGFRSTLTGKNYTWMEYFOLGNCETPHURPEMDAPFPCNOG 162
121  AACFEGIDVHMKENGTLVQVATISGMENQAKWKQKONETGIYYETWNVKASPEKGA 180
163  AACFEGIDVHMKENGTLVQVATISGMENQAKWKQKONETGIYYETWNVKASPEKGA 222
181  ETWDSYDCSKFVLRTNKLAEFGAEFKNIETNTRIFLVSGETTYGNETSVGPTGNK 240
223  ETWDSYDCSKFVLRTNKLAEFGAEFKNIETNTRIFLVSGETTYGNETSVGPTGNK 282
241  TLGLAIRFYPPKPHLPYKEFLISLQIFDAVIVHKQFYLFINFEYFELPMKFPPIKIT 300
283  TLGLAIRFYPPKPHLPYKEFLISLQIFDAVIVHKQFYLFINFEYFELPMKFPPIKIT 342
301  YEETPLPPIRNKTLISGL 316
343  YEETPLPPIRNKTLISGL 358

```

RESULT 4
 AAM19604
 ID AAM19604 standard; protein; 163 AA.
 XX
 XX
 AC AAM19604;
 XX
 XX
 DT 12-OCT-2001 (first entry)
 DE Peptide #6038 encoded by probe for measuring cervical gene expression.
 XX
 XX Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 XX
 OS Homo sapiens.
 XX
 EN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PD 30-JAN-2001; 2001WO-US000670.
 PF
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-488901/53.

Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human cervical epithelial cells.

Claim 27; SEQ ID NO 24430; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes
 CC (SENPs; see AAM10068-AAM128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at Ftp.wipo.int/pub/published_pct_sequences
 XX

Sequence 163 AA;

Query Match 50.4%; Score 886; DB 4; Length 163;

Best local Similarity 100.0%; Pred. No. 2.8e-81;

Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 154 AKWKQDNETGIYYETWNVKASPEKGAETWDSYDCSKFVLRTNKLAEFGAEFKNIETN 213
DB 1 AKWKQDNETGIYYETWNVKASPEKGAETWDSYDCSKFVLRTNKLAEFGAEFKNIETN 60
QY 214 YTRIFLYSGEPYTLGNETSVEGPTGNKTLGLAIRFYPPKPHLPYKEFLISLQIFDAV 273
DB 61 YTRIFLYSGEPYTLGNETSVEGPTGNKTLGLAIRFYPPKPHLPYKEFLISLQIFDAV 120
QY 274 IYHKQFYLFINFEYFELPMKFPPIKITYERIPPIRNKTLISGL 316
DB 121 IYHKQFYLFINFEYFELPMKFPPIKITYERIPPIRNKTLISGL 163

```

RESULT 5

BB39239
D ABB39239 standard; peptide, 163 AA.
X
X ABB39239;
X
T 04-FEB-2002 (first entry)
X
X Peptide #6745 encoded by human foetal liver single exon probe.
X
X Human; foetal liver; gene expression; single exon nucleic acid probe.
S Homo sapiens.
N WO200157272-A2.
D
X 09-AUG-2001.
X
X 30-JAN-2001; 2001WO-US000669.
X
X 04-FEB-2000; 2000US-0180312P.
X 26-MAY-2000; 2000US-0207456P.
X 30-JUN-2000; 2000US-00608408.
X 03-AUG-2000; 2000US-00632366.
X 21-SEP-2000; 2000US-0234687P.
X 27-SEP-2000; 2000US-0236359P.
X 04-OCT-2000; 2000GB-00024263.
X
X (MOLE-) MOLECULAR DYNAMICS INC.
X
X Penn SG, Hanzel DK, Chen W, Rank DR;
X WPI; 2001-483447/52.
X
X Human genome-derived single exon nucleic acid probes useful for analyzing
T gene expression in human fetal liver.
T
S Claim 27; SEQ ID NO 31874; 639pp + Sequence listing; English.
X
C The invention relates to a single exon nucleic acid probe for measuring
C human gene expression in a sample derived from human foetal liver. The
C single exon nucleic acid probes may be used for predicting, measuring and
C displaying gene expression in samples derived from human fetal liver. The
C present sequence is a peptide encoded by a single exon nucleic acid probe
C part of the invention. Note: The sequence data for this patent did not form
C directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
X
Q Sequence 163 AA;
Query Match 50.4%; Score 886; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 2.8e-81;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 154 AKWVKODNETGIYETWNVKASPEKGAETWFDSDSKFVLRFTFNKLAEPGAEFNKIEIN 213
b 1 AKWVKODNETGIYETWNVKASPEKGAETWFDSDSKFVLRFTFNKLAEPGAEFNKIEIN 60
Y 214 YTRIFLYSGEPYTLGNETSVPPTGNKKTGLAIKRFYFPKPHLPKKEFLSLLOTFDAV 273
b 61 YTRIFLYSGEPYTLGNETSVPPTGNKKTGLAIKRFYFPKPHLPKKEFLSLLOTFDAV 123
Y 274 IVHKQFYLFPNFEYWFLLPMKFPFIKITVEIPLPIRNKTLISGL 316
b 121 IVHKQFYLFPNFEYWFLLPMKFPFIKITVEIPLPIRNKTLISGL 163
RESULT 6
D ABB32740 standard; protein, 163 AA.
X
X AAB32740;
X
X 17-OCT-2001 (first entry)
T

XX
DE Peptide #6777 encoded by probe for measuring placental gene expression.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XW genetic disorder.
XX
OS Homo sapiens.
X
X WO200157272-A2.
X
X 09-AUG-2001.
X
X 30-JAN-2001; 2001WO-US000663.
X
X 04-FEB-2000; 2000US-0180312P.
X 26-MAY-2000; 2000US-0207456P.
X 30-JUN-2000; 2000US-00608408.
X 03-AUG-2000; 2000US-00632366.
X 21-SEP-2000; 2000US-0234687P.
X 27-SEP-2000; 2000US-0236359P.
X 04-OCT-2000; 2000GB-00024263.
X
X (MOLE-) MOLECULAR DYNAMICS INC.
X
X Penn SG, Hanzel DK, Chen W, Rank DR;
X WPI; 2001-48897/53.
X
X Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human placenta.
PT
X Claim 27; SEQ ID NO 33009; 654pp; English.
X
X The present invention relates to single exon nucleic acid probes (SENP;
CC see AAI31315-AI157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders
X
Q Sequence 163 AA;
Query Match 50.4%; Score 886; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 2.8e-81;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 154 AKWVKODNETGIYETWNVKASPEKGAETWFDSDSKFVLRFTFNKLAEPGAEFNKIEIN 213
Db 1 AKWVKODNETGIYETWNVKASPEKGAETWFDSDSKFVLRFTFNKLAEPGAEFNKIEIN 60
QY 214 YTRIFLYSGEPYTLGNETSVPPTGNKKTGLAIKRFYFPKPHLPKKEFLSLLOTFDAV 273
Db 61 YTRIFLYSGEPYTLGNETSVPPTGNKKTGLAIKRFYFPKPHLPKKEFLSLLOTFDAV 120
QY 274 IVHKQFYLFPNFEYWFLLPMKFPFIKITVEIPLPIRNKTLISGL 316
Db 121 IVHKQFYLFPNFEYWFLLPMKFPFIKITVEIPLPIRNKTLISGL 163
RESULT 7
D ABB24080 standard; protein, 163 AA.
X
X ABB24080;
X
X 23-JAN-2002 (first entry)
X
X Protein #6079 encoded by probe for measuring heart cell gene expression.
DE Human; gene expression; heart; microarray; vascular system;
XW cardiovascular disease; hypertension; cardiac arrhythmia;
XW congenital heart disease.
X

Homo sapiens.

XX
XX
30 JAN 2007 0000 UTC

30-0441-0500088 /
Y002; Y001;
XX
XX

PR		04-OCT-2000; 2000GH-00024263.	
XX	(MOLE-) MOLECULAR DYNAMICS INC.		
PA	Penn Sq, Hanzel DK, Chen W, Rank DR;		
PI	WPI, 2001-48898/53.		
XX			
DR			
XX			
PT	Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human adult liver.		
PS	Claim 27; SEQ ID NO 32831; 658bp; English.		
CC	The invention relates to a single exon nucleic acid probe (SENP) (I) for measuring human gene expression in a sample derived from human adult liver, comprising one of 13109 defined nucleotide sequences given in the specification (or complements/ fragments). The probe hybridises at high stringency to a nucleic acid molecule expressed in the human adult liver. (II) may be used for predicting, measuring and displaying gene expression in samples derived from human adult liver. The genes identified may be involved in genetic liver diseases such as cirrhosis,		
CC	hyperlipoproteinemia, hyperlipedema and hypercholesterolaemia which is associated with coronary heart disease. ABG47348-ABG59930 represent human liver single exon encoded peptides of the invention. Note: The sequence information for this patent does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences		
CC			
CC			
SQ	Sequence 163 AA;		
	Query Match Best Local Similarity 100.0%; Pred. No. 2.8e-81; Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0		
OY	154 AKWKQDNDEGTGYETNWNKASPEKGATWPDSYCCKFVLTEFNKAIEFGAEFKNIEN 213		
Dd	1 AKWKQDNETGIYYETNMVKNASPEKGAETWDSDSCKFVLTFNKLAIFGAFPKNIEN 60		
OY	214 YTRTFLVSGETTYLGNTSVSGFTGNKTLLGLAIRFYYPFPFHLPVKRPLLILLOITPDAY 273		
Dd	61 YTRTFIVSGETTYLGNSTVFSGPIONTKTLGLAIAIRFYYPFPFHLPVKREPLLSILOIPDAY 120		
OY	274 IVHKQFIYLFNFPEYFWLPMPKEPFIKITYEELPLPIRNKLTSL 316		
Dd	121 IVHKQFILFFNYEFWFLEMKPPFIKITYEELPLPIRNKLTSL 163		
RESULT 11			
ID	ABG42312 standard; peptide; 163 AA.		
XX	ABG42312;		
AC			
DT	19-AUG-2002 (first entry)		
DB	Human peptide encoded by genome-derived single exon probe SEQ ID 31977.		
XX			
KM	Human; single exon probe; asthma; lung cancer; COPD; ILD;		
KM	chronic obstructive pulmonary disease; interstitial lung disease;		
KM	familial idiopathic pulmonary fibrosis; neurofibromatosis;		
KM	tuberos scleriosis; Gaucher's disease; Niemann-Pick disease;		
KM	Hermansky-Rudrak syndrome; sarcoidosis; pituitary haemorrhage;		
KM	pulmonary histiocytosis; lymphangiomyomatosis; Karsenger syndrome;		
KM	pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;		
KM	primary ciliary dyskinesia; pulmonary hypertension;		
XX	hyaline membrane disease.		
OS	Homo sapiens.		
XX			
FN	WO200186003-A2.		
XX			
ID	15-NOV-2001.		

30-JAN-2001; 2001WO-US000665.
30-FEB-2000; 2000US-0180312P.
26-MAY-2000; 2000US-0207456P.
30-JUN-2000; 2000US-00608408.
03-AUG-2000; 2000US-00632366.
21-SEP-2000; 2000US-0234687P.
27-SEP-2000; 2000US-0236359P.
04-OCT-2000; 2000GB-00024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2002-114183/15.
Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples.
Claim 27; SEQ ID NO 31977; 634bp; English.
The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsageners syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences
Sequence 163 AA:
Query Match 50.4%; Score 886; DB 5; Length 163;
Best Local Similarity 100.0%; Pred. No. 2.8e-81;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
154 AKWVKQDMEGTGYETWVWVKASPEKGARTWPSDYDCSEVLRITNKIAEFGAEFNIETN 2.3
Db 1 AKWVKQDMEGTGYETWVWVKASPEKGARTWPSDYDCSEVLRITNKIAEFGAEFNIETN 60
214 YTRIFLVSGEPTLYGNETSVEFPTGNKTLGAIKRFYFFXPHLPTEKFLLSLQIFDVA 273
Db 61 YTRIFLVSGEPTLYGNETSVEFPTGNKTLGAIKRFYFFXPHLPTEKFLLSLQIFDVA 120

274 IYHKQFLYFNFEYXWFLPMKEPFIKTYEIPRIRKUTLSGL 316
Db 121 IYHKQFLYFNFEYXWFLPMKEPFIKTYEIPRIRKUTLSGL 163
RESULT 12
ABP42902
ID ABP42902 standard; protein; 89 AA.
XX
AC ABP42902;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HRPDU81, SEQ ID NO:4034.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive.
XX
OS Homo sapiens.
XX
PN MO200200677-A1.
XX
PD 03-JAN-2002.
PP 07-JUN-2001; 2001WO-US018569.
PR 07-JUN-2000; 2000US-0209467P.
PS (HUMA-) HUMAN GENOME SCI INC.
PI Birse CE, Rosen CA;
XX
XX WPI; 2002-147878/19.
DR N-PSDB; ABQ55979.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
PT cancer), immune disorders, cardiovascular disorders and neurological
PS
PS Claim 11; SEQ ID NO 4034; 2922bp; English.
XX
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
XX ABP41328) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
XX encompasses polypeptides 90% identical and polynucleotides 95% identical
XX to the sequences of the invention. The invention additionally relates to
XX recombinant vectors and host cells comprising human ovarian antigen
XX polynucleotides, antibodies against human ovarian antigens, and the use
XX of ovarian antigen polynucleotides and polypeptides in diagnosing,
XX treating, prognosing or preventing various ovarian and/or breast-related
XX disorders. Such conditions include ovarian cancer and breast cancer, and
XX metastatic tumours of ovarian or breast origin, reproductive system
XX disorders (e.g., infertility, disorders of pregnancy, anovulation,
XX polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
XX disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
XX shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
XX vaginitis), immune disorders (e.g., congenital and acquired
XX immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
XX blood-related disorders (e.g., anaemia), cardiovascular disorders,
XX respiratory disorders, neurological disorders, gastrointestinal disorders
XX and urinary system disorders. Ovarian antigen polypeptides and
XX polynucleotides may also be used in screening for compounds which
XX modulate ovarian antigen expression or activity. The polynucleotides may
XX further be used for gene therapy, chromosome mapping, in the

CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 89 AA;

Query Match 26.8%; Score 472; DB 5; Length 89;
Best Local Similarity 98.9%; Pred. No. 1e-39;
Matches 88; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2Y 228 GNETSVFPGTGNKTLGLAIKRFYFPKPHLPTRKFLSLQLFDPAVIVHKQFYLRYNFEY 237
DB 1 GNETSVFPGTGNKTLGLAIKRFYFPKPHLPTRKFLSLQLFDPAVIVHKQFYLRYNFEY 63
2Y 288 WFLPMKFPFIKITYPEIFLPTRNKTLSQL 316
DB 61 WFLPMKFPFIKITYPEIFLPTRNKTLSQL 89

RESULT 13

AB60074 ID AB60074 standard; protein; 2703 AA.

AC AB60074;

2T 26-MAR-2002 (first entry)

2E Drosophila melanogaster polypeptide SEQ ID NO 7014.

2W Drosophila developmental biology; cell signalling; insecticide;
2X pharmaceutical.

2S Drosophila melanogaster.

2N WO200171042-A2.

2D 27-SEP-2001.

2F 23-MAR-2001; 2001WO-US009231.

2R 23-MAR-2000; 2000US-0191637P.

2R 11-JUL-2000; 2000US-00614150.

2A (PEKE) PE CORP NY.

2I Venter JC, Adams M, Li PWD, Myers EW;

2R WPI; 2001-656860/75.

2R N-PSDB; ABL04177.

2T New isolated nucleic acid detection reagent for detecting 1000 or more
2T genes from Drosophila and for elucidating cell signaling and cell-cell
2T interactions.

2S Disclosure; SEQ ID NO 7014; 21pp + Sequence Listing; English.

2C The invention relates to an isolated nucleic acid detection reagent
2C capable of detecting 1000 or more genes from Drosophila. The invention is
2C useful in developmental biology and in elucidating cell signaling and
2C cell-cell interactions in higher eukaryotes for the development of
2C insecticides, therapeutics and pharmaceutical drugs. The invention
2C discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
2C sequences (AB101840-AB116175) and the encoded proteins (AB57737-
2C AB572072). The sequence data for this patent did not form part of the
2C printed specification, but was obtained in electronic format directly
2C from WIPO at ftp.wipo.int/pub/published_pct_sequences

2X Sequence 2703 AA;

Query Match 5.4%; Score 95.5; DB 4; Length 2703;
Best Local Similarity 20.1%; Pred. No. 21;
Matches 71; Conservative 49; Mismatches 139; Indels 95; Gaps 17;

QY 1 RVSGIPSRHMPVYKREDFPKDPYQAKYTCPTGSPFVPMGDDIEVRLQAPVW 60
DB 1966 RFNKIPQSEW-----KSKRPHPELWKKLEKT-----PYKGGNSLRPYCLEGLNW 2032

QY 61 EFYKGDLLGHLKIMHDAIGFASSTLTGNKYTWMEYELFQAGCTPHLRPEMDAPWCNQS 120

DB 2033 -LKSFWYNHNCILADEWGLAKTIQSLTFHVSVEYGRG--PLVIAPIASTIPNMQRE- 2088

QY 121 AACFEEDIDVHMKENGTLVQVATISGNMFQMAKVMQDNETGIYETMVKASPE--- 177

DB 2089 ---FEG-----WIDMNVVYHGSVT-----SKQMIDYE--YYITBEGKYKEIK 2130

QY 178 -KGAETWDSYDCKSFVLRTN-----KLAEGAEKNTE----- 211

DB 2131 FNVLLITTFEMITVDYMDLKAENWRLCVIDEARLKNRCKLLE-GLROLNIEHRYLLSGT 2189

QY 212 ---TNYRIF--LYSGEPTVIGNETSVFGPTGNKTLGLAIKRFYFPKPHL----- 257

DB 2190 PLQNNISELPSLNFLEBSPSSQSEFMSEFSGSLRTEBEVVKLQALLKPMMLRLKLDVE 2249

QY 258 ---PTKEFLSLQLIPDAVIVHKQFYLRYNFEYFPLPMKPPFIK--ITYEIP 305

DB 2250 KSLAPKETEITIEV---ELTNIQKKY-----RGILEQNSFLKKGTSANIP 2293

RESULT 14

ABU52308 ID ABU52308 standard; protein; 429 AA.

AC ABU52308;

2T 07-MAY-2003 (first entry)

2E Helicobacter pylori selected interacting domain (SID) protein #1651.

2W Helicobacter pylori selected interacting domain (SID) protein #1651.

2S Helicobacter pylori.

2N WO200266501-A2.

2D 29-AUG-2002.

2F 28-DEC-2001; 2001WO-EP015428.

2R 02-JAN-2001; 2001US-0259302P.

2A (HYBR-) HYBRIGENICS.

2I (INSP) INST PASTEUR.

2T Legrain P, Rain J, Colland F, De Reuse H, Labigne A;

2R WPI; 2002-674910/72.

2R N-PSDB; ABX70383.

2T New complexes of protein-protein interactions in Helicobacter pylori,
2T useful for identifying modulating compounds for treating or preventing
2T ulcers in mammals.

2S Claim 6; Page 618; 642pp; English.

2C The invention describes a complex of protein-protein interactions in
2C Helicobacter pylori selected from 421 complexes given in the
2C specification. The complex of protein-protein interactions are useful for
2C screening for agents which modulate the interaction of proteins.
2C Modulating compounds which binds to a targeted bacterial protein may be
2C used for treating or preventing ulcers in a human or animal. This is the
2C amino acid sequence of a selected interacting domain (SID), identified
2C via protein-protein interactions. Note: Where the patent number printed

at the top of the pages in the specification has obscured areas of protein sequence, the indexer has replaced the residue with an X to represent an illegible residue

Sequence 429 AA:

Query Match	Score	DB	Length
5.3%	94	5	429

Local Similarity 23.6%; Pred. No. 2.1,
62: Conservative 28: Mismatches

atches 62; Conservative 28; Mismatches 77; Indels 96; Gaps 15;

```

80  FRSLLTKKNTVMWEYVEFQIAGNCFPHLRPEMDAPFCN-----QGAACEFEGIDVHYMK 134
125  FTQYGGKNSALVY-----NATLP-----MANGSIKSNSTVAFGIGYEGVNWG 166
135  ENGLTVQVAT-----ISGNFNMQAKVWKQDNQETGIYETWNVKAPEKGAETWFDSDYC 184
168  KTGIIITGTFTADRYVLIINGNM-----TG-----NSAQITGGGA--T 200
190  SKFLARFENKLAERGAERKNIET---NYTRIELASGEPTVLGNETSVFGTQNKITGLAI 244
201  LNFVGATEINIA--GAFTPKDKIKTSSONSNTFMAIAG-----SSGSAKINVSQ 244
247  KRFY-----YPEKPHLPTEKEFLSLLIQIFDAVAVKQFYLF-----YFEEYV-FLPMK 293
247  SDFEDWTGGGQYDFNG-----GVFDSVVENKAYKFKQSTENSYNFKNTNIFLAGN 295
294  PPFY-KITYEIPLPINKKTLTG 315
296  FKFGKTTIKESVLDASSTYFDG 318

```

SUIT 15

ABU30825 standard; protein; 1238 AA.

ABU30825;

19-JUN-2003 (first entry)

Protein encoded by Prokaryotic essential gene #16352.

Antisense; prokaryotic essential gene; cell proliferation; drug design.

Helicobacter pylori.

MO200277183-A2

03-OCT-2002.

21-MAR-2002; 2002WO-US0091.07

21-MAR-2001; 2001US-00815242

25-OCT-2001; 2001US-0342923P
08-FEB-2003; 2003US-00073851

06-MAR-2002; 2002US-0362699P.

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C

WPT: 2003-039926/02

N-PSDB; ACA34695.

New antisense nucleic acids, useful for homologous nucleic acids

Isolate candidate mo

Claim 25; SEQ ID NO 58749; 1766bp; English.

the 6213 antisense sequences given in the specification where expression of the invention relates to an isolated nucleic acid comprising any one of

CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activation of a gene in an operon required for
 CC proliferation; (7) identifying a gene in an operon required for
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-regulated gene or its gene product lies
 CC or a gene on which the test compound directly inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

Sequence 1238 AA;

Query Match	Score	DB	Length
5.3%	94	6	1238

Best Local Similarity 23.6%; Pred. No. 9.77
Matches 63; Conservative 38; Wismatches 1

Matches	62;	conservative	28;	Mismatches	77;	Indels	96;	Gaps	15;
---------	-----	--------------	-----	------------	-----	--------	-----	------	-----

QY 80 FSTLLGKAYTMEWELEFQNGCTPPHRLPEMDAPFCN-----QGAACFEEDIDYHK 134
Db 513 FQGTGCKRKSALVF-----NATLP-----WANGSIPIPSNINVRGGIIEGVNMG 555
QY 135 ENGLTVOVAT-----ISGNMENQAKWVKQDNETGIIYYETWNVKASPEKGAETWFDSDYC 189
Db 556 KKGYLITGETADRVAYITANM-----TG-----NGAQITGGGA--T 588
QY 190 SKFVLRTFNKLAEFAEPRNLET---NYTRIFLTSGEPTLYGNETSVPGRGNKTLGLAI 246
Db 589 LNFVGATELNID--GATRNKLKTSQNSYMTFMALGD-----SSGASKXIVSQ 634
QY 247 KRFY-----YFKPHLPTEKFLSLQIPDAVIVHQFYLF-----YNEFYW-FLPMK 293
Db 635 SDFYMTGAGGYDFTGN-----GVFDSVNFNKAYYKFGCTENSYNFKNTNPLAGN 683
QY 294 FPEL-KITYEELPDIRKKTLSG 315
Db 684 FFKGKTLIEKSVLSDASYTFDG 706

Search completed: June 8, 2004, 14:03:58
Job time : 84.8278 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

l protein - protein search, using sw model

on: June 8, 2004, 14:00:33 ; Search time 24.5778 seconds

(without alignments)
663.762 Million cell updates/sec

file: US-10-010-050A-2_COPY_31_346

sequence: 1 RVSGSPRRHWPVYKRFDF.....IKITYEIPLPFRKNTLSGL 316

oring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

minimum DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

tabase :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backkilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	91	5.2	4563	4	US-09-108-006C-1	Sequence 1, Appli
2	88.5	5.0	1462	3	US-07-792-600-31	Sequence 31, Appl
3	88.5	5.0	1462	3	US-09-157-021-31	Sequence 31, Appl
4	88.5	5.0	1462	3	US-09-156-842-31	Sequence 31, Appl
5	88.5	5.0	1462	4	US-09-591-514-31	Sequence 31, Appl
6	88	5.0	608	4	US-09-489-039A-13503	Sequence 13503, A
7	87.5	5.0	178	4	US-09-134-000C-3624	Sequence 3624, Ap
8	83	4.7	535	4	US-09-134-078-24	Sequence 24, Appl
9	83	4.7	1056	4	US-09-079-030-217	Sequence 217, App
10	83	4.7	4536	4	US-09-180-422B-27	Sequence 1, Appli
11	83	4.7	4536	4	US-09-079-030-1	Sequence 1, Appli
12	82	4.7	876	4	US-09-254-352B-19	Sequence 19, Appl
13	81	4.6	741	4	US-09-488-039A-7854	Sequence 7854, Ap
14	80.5	4.6	718	4	US-09-328-352-4640	Sequence 4640, Ap
15	80	4.6	451	4	US-09-489-039A-11849	Sequence 11849, A
16	80	4.6	607	4	US-09-537-682-1	Sequence 1, Appli
17	80	4.6	652	4	US-09-489-039A-12461	Sequence 12461, A
18	79.5	4.5	197	4	US-09-489-039A-8317	Sequence 8317, Ap
19	78.5	4.5	634	4	US-09-849-334-4	Sequence 4, Appli
20	78.5	4.5	634	4	US-10-274-878-4	Sequence 4, Appli
21	78.5	4.5	669	4	US-09-134-000C-6185	Sequence 6185, Ap
22	77	4.4	238	3	US-08-634-475-6	Sequence 6, Appli
23	77	4.4	238	3	US-08-709-791-6	Sequence 6, Appli
24	77	4.4	405	4	US-09-134-001C-5259	Sequence 5259, Ap
25	77	4.4	501	2	US-08-781-802-4	Sequence 4, Appli
26	77	4.4	501	2	US-08-781-802-6	Sequence 6, Appli
27	77	4.4	501	3	US-08-694-078-4	Sequence 4, Appli

28	77	4.4	501	3	US-08-694-078-6	Sequence 6, Appli
29	77	4.4	501	3	US-09-058-260-4	Sequence 4, Appli
30	77	4.4	501	3	US-09-058-260-6	Sequence 6, Appli
31	77	4.4	501	3	US-09-058-260-14	Sequence 14, Appli
32	77	4.4	501	3	US-09-058-260-20	Sequence 20, Appli
33	77	4.4	501	3	US-09-058-260-22	Sequence 22, Appli
34	77	4.4	501	3	US-09-058-260-24	Sequence 24, Appli
35	77	4.4	501	3	US-09-058-260-32	Sequence 32, Appli
36	77	4.4	615	4	US-09-107-532A-6507	Sequence 6507, Ap
37	77	4.4	615	4	US-09-107-532A-4897	Sequence 4897, Ap
38	76	4.3	676	4	US-09-107-532A-5806	Sequence 5806, Ap
39	76	4.3	537	4	US-09-462-284-2	Sequence 2, Appli
40	75.5	4.3	537	4	US-09-489-039A-14149	Sequence 14149, A
41	75.5	4.3	707	1	US-08-528-122-18	Sequence 18, Appl
42	75.5	4.3	707	5	PCT-US95-11720-18	Sequence 18, Appl
43	75.5	4.3	750	1	US-08-325-553-2	Sequence 2, Appli
44	75.5	4.3	750	1	US-08-394-152A-2	Sequence 2, Appli
45	75.5	4.3	750	3	US-09-044-668-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-108-006C-1
Sequence 1, Application US/09108006C
Patent No. 6524613

GENERAL INFORMATION:

APPLICANT: Steer, Clifford J.

Kren, Betsy T.

Bandyopadhyay, Paramita

Roy-Chowdhury, Jayanta

TITLE OF INVENTION: Hepatocellular Chimeraplasty

NUMBER OF SEQUENCES: 62

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kimeragen, Inc.

STREET: 300 Pleasant Run

CITY: Newtown

STATE: PA

COUNTRY: USA

ZIP: 18940

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/108,006C

FILING DATE: 30-Jun-1992

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/054,288

FILING DATE: 30-APR-1997

APPLICATION NUMBER: 60/054,837

FILING DATE: 05-AUG-1997

APPLICATION NUMBER: 60/064,996

FILING DATE: 10-NOV-1997

APPLICATION NUMBER: 60/074,497

FILING DATE: 12-FEB-1998

APPLICATION NUMBER: PCT US 98/08834

FILING DATE: 30-APR-1998

ATTORNEY/AGENT INFORMATION:

NAME: Friebe, Thomas

REGISTRATION NUMBER: 29258

REFERENCE/DOCKET NUMBER: 7991-015-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-504-4444

TELEFAX: 215-504-4545

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4563 amino acids

TYPE: amino acid

SEQ ID NO 31

[illegible]

Query Match	5.0%;	Score 88;	DB 4;	Length 608;
Best Local Similarity	20.9%;	Pred. No. 0.64;		
Matches	67;	Conservative	41;	Mismatches 95; Indels 118; Gaps 21;

QY	49	DIEVRLQAPWEFKYGDLLGHLKIMDA--TGRSTLTGK-NYTMF--WYELPOLGNC	102
		: : : : : : :	
Db	215	DTEDYRRVDP---QFGDALLRLRHNTORAGRWMLDDGFVNTGDSHPEDFRHQOQSG	270
		: : : : : : :	
QY	103	TFPLRLREMDPF--WNCQGAAC-----FFSGIDVV--HWK	134
		: : : : : : :	
Db	271	GAGH---DPDSFWRDMFTFSEEGAHNMLGYASLPKLDYRSTSLVNEIYAEBSIDVRHML	327
		: : : : : : :	
QY	135	E-----NGTLVQVATIS-----GNMFMQAKWVKQ	159
		: : : : : : :	
Db	328	KAPSMQNGWRUDVYHMLGEGSGARANNIOLHLAGITQAKQAQPEAFVSEHHTGDARQWLQA	387
		: : : : : : :	
QY	160	D-----NEGSIYETW-----NFKASPEK-GAET--WFDSDCS--KFLVRLTFKL	200
		: : : : : : :	
Db	388	DAEDAMNRYRFTTPFINGFLANDISYDPOIDAQOTCMAMWMDNRYRAGISHQOOLRMFQOL	447
		: : : : : : :	

Y 201 -----ARF-----GAFFKNIETNYTRIFLYSGEP-TYLGNETSVFG---PTGNKTLGLAIK 247
b 448 DSHDLARFKSLLDKDVARDLAVVWLFSGVGVCIYIGDEVGDANNDFPCRRP----- 501
Y 248 RRYPPKPLPTKEFLISLIQ 268
b 502 ---FPWDFALQDTQ-LIALYQ 518

RESULT 7

S-09-134-000C-3624

Sequence 3624, Application US/09134000C

Patent No. 6617156

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 032796-032

CURRENT APPLICATION NUMBER: US/09/134,000C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/055,778

PRIOR FILING DATE: 1997-08-15

NUMBER OF SEQ ID NOS: 6812

SOFTWARE: PatentIn version 3.1

SEQ ID NO 3624

LENGTH: 178

TYPE: PRT

ORGANISM: Enterococcus faecalis

S-09-134-000C-3624

Query Match 5.0%; Score 87.5; DB 4; Length 178;
Best local Similarity 21.6%; Pred. No. 0.11;

Matches 41; Conservative 24; Mismatches 60; Indels 65; Gaps 9;

Y 87 KNTTMEWYELFQIGCTFPHLRPEMDAPFMCNCGAACFFEGIDDVHMKENGTLVQVATIS 146
b 13 KNTTRELGETDLSKGYISQLEHDLSP-----SMETFF-----SILEVLGYT 55
Y 147 GNNENQM-----AKWQNDNETGIYE-----TWNVKASPEKGATFWDPSYCSKFLVR 195
b 56 PEGFPGQPLEQKIVINEESTLYDDEHGYELKMLIPASNEKEMEP-----VII 105
Y 196 TENKLAFFGAFFNIETNYTRIFLY-----SGEPFLGNETSVFGPTGNKTLGLAI 245
b 106 TEKK-----NKEYTFEPSSLSEFTIYVIDGSIRTLGEATYIAKK-----GQT----- 143
Y 247 KRFYPPKPR 256
b 149 -MYQATEPH 157

RESULT 8

S-09-134-078-24

Sequence 24, Application US/09134078

Patent No. 6368844

GENERAL INFORMATION:

APPLICANT: Bvllina, Edward J.

TITLE OF INVENTION: GLYCOSIDASE ENZYMES

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Gray Cary Ware & Freidenrich LLP

STREET: 4365 Executive Drive, Suite 1600

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/134,078
FILING DATE: 13-AUG-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/949,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/024002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858/677-1456
TELEFAX: 858/677-1465
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 555 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-134-078-24

Query Match 4.7%; Score 83; DB 4; Length 555;
Best local Similarity 22.1%; Pred. No. 2.1;
Matches 50; Conservative 32; Mismatches 68; Indels 76; Gaps 12;

QY 34 FCPGSPPIVMEGDDIEVRLQAPVWERKYGDLLKIMHDAIFRSSTLTCKNTMEX 93
b 149 FVPL-EPVLVLEDPN-----TPLLERKVALVG---MENNARVPKHPTG---WCSW 193
QY 94 YELF--QLGNCFTPHLRPEMDAPFMCNCGAACFFEGIDDVHMK----- 134
b 194 YHFFDLTWEEFLKIKLAKNPFVYQ-----IDAYEDIGDWLVTGDFPSVBE 245
QY 135 -----ENGLVQVATI-----SGNMENQAKW-VKQDNETGIYETWNVKASPEKGAB 181
b 246 MAKVIAENGFIPGIMTAPPSVSETSDVFNHEDVYKENGELKMAVRYNNKK----- 297
QY 182 TWFDSDYDCK-----FVLRTEKKLAFFGAFFNIETNYTRIFLYSG 222
b 298 --IYALDISKDEVLNWLPDLFSLRKMGRYRFXID-----PLFAG 345

RESULT 9

US-09-079-030-217

Sequence 217, Application US/09079030

Patent No. 6635623

GENERAL INFORMATION:

APPLICANT: Guevera, Jr., Juan G.

APPLICANT: Hoogeveen, Ron C.

TITLE OF INVENTION: LIPOPROTEINS AS NUCLEIC ACID DELIVERY

TITLE OF INVENTION: VECTORS FOR TRANSFECTION OF EUKARYOTIC CELLS

NUMBER OF SEQUENCES: 229

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/079,030

FILING DATE: Concurrently Herewith

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

TOPOLOGY: linear
S-09-079-030-1

Query Match 4.7%; Score 83; DB 4; Length 4536;

Best Local Similarity 19.7%; Pred. No. 54; Matches 56; Conservative 39; Mismatches 123; Indels 66; Gaps 11;

Y 44 MEGDDDEVERLQAPWEFKYGDLLG---HLKIMHDAIGFRSTLTGKNYTWETELFQLG 100
b 3986 MDEDDDS-----KNFYYSPSSDPDKLTTFKTELRRSEDEEIQIVNNEEELASG 4038
Y 101 NCTFPHLRPEMDAPFWCNGAACEFEGIDVHWKENG-TLVQVATISGNFNQMAKWKQ 159
b 4039 LLT--SIKQVNP-----KATGVLYDVYKNKXHEHTGLTTRVSSSKRRRLQNNAEWVYQ 4090
Y 160 DNEGIYETWNTVYASPEKAEETWEDSYDSKYVLRTRFNKLAERGAEPKIEFN-YTRIF 218
b 4091 ----GAIRQIDDDIVRFQKASGTTGYQ-----EMKQKQNLVQEL 4129
Y 219 LYSGEPTYLGNETSVEFGPTGNKTLGLAIK-----REYYPEKPHLPYKEFL 263
b 4130 TQEGQASFOGLKONVFGIVRVQKFMKXVHLIDSLIDFLNFPFPQFPQKPGIYTRREL 4189
Y 264 LSLIQIPDAVIYHKQFY-----LFYNEYWFLPEKPEPTK 298
b 4190 CTWF-IREVGTVLQVYSKYHNGSEILFSYFODLIVITLPELBR 4232

RESULT 12
S-09-254-352B-19

Sequence 19, Application US/09254352B
Patent No. 6365350

GENERAL INFORMATION:
APPLICANT: HAYASHIZAKI, Yoshihide

TITLE OF INVENTION: METHOD OF DNA SEQUENCING

FILE REFERENCE: 024705-080

CURRENT APPLICATION NUMBER: US/09/254,352B

CURRENT FILING DATE: 1999-11-10

PRIOR APPLICATION NUMBER: PCT/JP98/03039

PRIOR FILING DATE: 1998-07-06

PRIOR APPLICATION NUMBER: JP 10-155847

PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: JP 9-196478

PRIOR FILING DATE: 1997-07-07

NUMBER OF SEQ ID NOS: 64

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 19

LENGTH: 876

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: T7 RNA polymerase

S-09-254-352B-19

Query Match 4.7%; Score 82; DB 4; Length 876;

Best Local Similarity 22.3%; Pred. No. 5.5;

Matches 45; Conservative 22; Mismatches 67; Indels 68; Gaps 12;

Y 35 CPTGSPFPWEGD-----DDIEVRLQAPWEFKYDGLGLKIM----- 74
b 356 CPVGD-VPAIERELPRPDIDITNEVARAKMRKRAAAYVRKOKAROSRRCEFWAOA 414
Y 75 ----HDAIGRSTLTGKNYTWETELFQLG-----CTFPHLRP-ENDAPWCN 118
b 415 NFNANKALWF-----PYNDWRGRYVAVSMGNDKTSKSLTLAKGKYGIDGFWLTK 467
Y 119 -QGAACFEGIDVHWKENGTLVQVATISGNFNQMAK-----WYQDNTEGTIYYE 168
b 468 IHGANC-AQVDKVPPE-----RIKFIENEGNILLASADPLNNTWMTQODSPAPACFE 520
Y 169 TNNYKASPEKGAETWFPDSYCS 190
b 521 YAGVK-----HNGINSYCS 535

RESULT 13

US-09-489-039A-7854

Sequence 7854, Application US/09489039A

Patent No. 6610836

GENERAL INFORMATION:

APPLICANT: Gary Breton et. al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2004001

CURRENT APPLICATION NUMBER: US/09/489,039A

CURRENT FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: US 60/117,747

PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 14342

SEQ ID NO 7854

LENGTH: 741

TYPE: PRT

ORGANISM: Klebsiella pneumoniae

US-09-489-039A-7854

Query Match 4.6%; Score 81; DB 4; Length 741;

Best Local Similarity 22.7%; Pred. No. 5.5;

Matches 56; Conservative 31; Mismatches 66; Indels 94; Gaps 17;

Y 63 KYGDL-GHLKIMHDAIGFRSTLTG---KNYTWETELFQLGCTFPHLRPEMDAPF--- 115
b 518 KYGAAYNGHLPILSSVNDSPLOGIKVNSYELG-----RYTQDVLRTOIAAYYSIS 570
Y 116 -----WQNGAACFEGIDVHWKENGTLVQVATISGNFNQMA 153
b 571 DKTIVNRDMTIDVQSDKQRIYCVGAVDYF--IDSYW-----SVGGN-FNVL 617
Y 154 AKWYQDNTEGTIYEFWYNY-ASPEKGAETW-----PDSYCSKFWLR 195
b 618 KSQVQYDGR---WQKMDVTLASPSK-ATAWGMADPWSLRVQSOQVPLSDAAGNKL 672
Y 196 TENKLAEPGA-----EFKN-LEITNYTRIF-----LYSGEPTYLGNETSVEFGPT 238
b 673 GYNTVDFISYALPVGKLTFSIENTLIMEDYVITMGORAPLIS-PTY-GSSSLIYBYK 728
Y 239 -NKTGL 244
b 729 RGRTFGL 735

RESULT 14

US-09-328-352-4640

Sequence 4640, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03PA

CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 4640

LENGTH: 718

TYPE: PRT

ORGANISM: Acinetobacter baumannii

US-09-328-352-4640

Query Match 4.6%; Score 80.5; DB 4; Length 718;

Best Local Similarity 20.6%; Pred. No. 5.9;

Matches 66; Conservative 41; Mismatches 123; Indels 91; Gaps 16;

Y 25 DPYQAKYTFPCPTSPWEGD--DIEVRLQAPWV-----EFKYGD 66
b 194 DPTQY--TYLAEALYEMGTRKDAAGLDAIKLGLMSDIFVRNPNKINGAADQENTED 251

```
Y      67  LIGHKIMEDAL-GERSTLTG--KNTTMEK--YELFO-IGNCTPPLRPMDAPFCNQG 120
      252  SNGIISTSDASRANRLRLOGINNFSLNGSRLOVFTLGAT-----TSGG 298
Y      121  AACFFE---GIDVHWKENGTLVQVATISGMFNO---MAKWK----- 158
      299  MSFFNNNTTIGIACVIRLNSGDSKVAKAITSSLTGSGTSPWTLIHAGANSTLISTTGD 358
Y      159  -QDNETGI-----YYETWNVKASPEKGAETWFDSDCKFVLRTFNKLAFFGAEPKN 203
      359  CNGGGSFGTSGAGCYVEKRTKRTSKATKTWDASGLSNAGVRLSTR----- 403
Y      210  IETNYTRILYSG---EPTVIGNE-TSVGPTGNKTLGLAIKRPYPKPHLPTEPL 263
      409  -ETSDTGNLITPAINGVAPTEPDANEGVILYNPINILVIG---TLVQPLVIGSDGKNFS 463
Y      264  LSLLOTFDAVIVHKQPYLFYN 284
      464  LEIARIANKPEIYKQIYTDYS 484
```

RESULT 15

3-09-489-039A-11849

Sequence 11849, Application US/09489039A

Patent No. 6610836

GENERAL INFORMATION:

APPLICANT: Gary Breton et. al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2004001

CURRENT APPLICATION NUMBER: US/09/489, 039A

CURRENT FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: US 60/117,747

PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 14342

SEQ ID NO 11849

LENGTH: 451

TYPE: PR

ORGANISM: Klebsiella pneumoniae

3-09-489-039A-11849

Query Match

Best Local Similarity 4.6%; Score 80; DB 4; Length 451;

Matches 56; Conservative 27; Mismatches 86; Indels 90; Gaps 14;

```
40  PIPVVEGD---DDIEVFLQAPVMEFK-----YGDLL 68
      151  PLADMDGKNLDDAFAFFGQDDWKIKIGREFAWDMFPLNDPTLEYSGNTANDLYSDGY 210
      69  GHLKIMHDAIGFRSTLTGK-----NYTWMEYELFQJGNTCTPPLRPEMDAPFCNQG 120
      211  GYIYMWKKEGRG-RSSSGGNLLEFSTVDNMYFEVNTLVLEDGSLF-----VDQNYHGNA- 262
      121  AACFEEGIDDV-----HWKENGTLVQVATISGMFNO---MAKWVKQDNETGI- 165
      263  ---LNNKXNVVYVPVAAW-QSGAWSAAMAIENSLVNNAYGYOSQGRWVDSNRGTGYG 317
      166  YETWNV-KASPEKGA-----ETWFDSDCKSEVARTFNKLAFFGAEPKNIFTNTYTRIFL 213
      318  LTMGNNTILKSDPDQDAVNLSTALIDADETDF-----GAGINALMHRVELGY----- 365
      220  YSGEPTVIGNETSVGPTG 238
      366  ---LYAHNKLDQENMAG 379
```

Search completed: June 8, 2004, 14:09:17
db time : 25.5778 secs


```

RESULT 3
JS-10-653-595-126
; Sequence: 126, Application US/10653595
; Publication NO. US20040048304A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 95 Human secreted proteins

```

RESULT 4
US-09-397-945-126
; Sequence 126, Application US/09397945
; Publication No. US20030065139A1

GENERAL INFORMATION:

APPLICANT: Human Genome Sciences, Inc. et al.
TITLE OF INVENTION: 95 Human secreted proteins
FILE REFERENCE: P2027P1

CURRENT APPLICATION NUMBER: US/09/397,945

PRIOR FILING DATE: 1999-09-17

PRIOR APPLICATION NUMBER: PCT/US99/05804

PRIOR FILING DATE: 1999-03-18

PRIOR APPLICATION NUMBER: 60/078,566

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,576

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,573

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,574

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,579

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/080,314

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/080,312

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/078,578

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,581

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,577

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,563

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/080,313

PRIOR FILING DATE: 1998-04-01

NUMBER OF SEQ ID NOS: 470

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 126

LENGTH: 347

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (242)

NAME/KEY: SITE

LOCATION: (246)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (347)

OTHER INFORMATION: Xaa equals stop translation

3-09-397-945-126

Query Match

Best Local Similarity

Matches 314; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
1 RVSGIPSRHMPVVKRPDPKPPDYQQAAYTCPTGSPPIPVMEGDDDIIVFLLQAPVW 60
2 31 RVSGIPSRHMPVVKRPDPKPPDYQQAAYTCPTGSPPIPVMEGDDDIIVFLLQAPVW 90
61 EFKYGDLLGHLKIMHDAIGFRSTLTGKNYTWMEYELFOLGNCSTPHLRPEMDAFPCNOG 120
91 EFKYGDLLGHLKIMHDAIGFRSTLTGKNYTWMEYELFOLGNCSTPHLRPEMDAFPCNOG 150
121 AACFPEGIDDVHMKENGTLVYVATISGMFNOMAKWVQNDNETGIYETWNVKASPEKGA 180
151 AACFPEGIDDVHMKENGTLVYVATISGMFNOMAKWVQNDNETGIYETWNVKASPEKGA 210
181 EFWEDSYDCSFVLRTEFKLAEPGAEPFNIEETNYTRLFLVSGEPFLYGNETSVPFGTKN 240
211 EFWEDSYDCSFVLRTEFKLAEPGAEPFNIEETNYTRLFLVSGEPFLYGNETSVPFGTKN 270
241 TLGAIKRFYVPFKPHLPTEKELLSLQIPDAVTHKQFYLFPNFEYFWLPMKPEPKIT 300
271 TLGAIKRFYVPFKPHLPTEKELLSLQIPDAVTHKQFYLFPNFEYFWLPMKPEPKIT 330
```

QY 301 YEELPLIRNKTLISGL 316
Db 331 YEELPLIRNKTLISGL 346

RESULT 5

US-09-664-761-39378

Sequence 39378, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecolica-X-1

CURRENT APPLICATION NUMBER: US/09/664,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 39378

LENGTH: 163

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: MAP TO AC00126.1

LOCATION: (163)

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6

OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.9

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3

OTHER INFORMATION: EST_HUMAN HIT: H01255.1, EVALUATE 5.00e-59

```
; OTHER INFORMATION: SWISSPROT HIT: 075503, EVALUATE 2.00e-36
JS-09-864-761-39378
Query Match          50.4%; Score 886; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 2.7e-80;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 154 AKWVKQDNNGIYETNNVAVASPEKGAETWEDSYDCKFYLRTFNKLAERGAFFKNIETN 213
    1 AKWVKQDNNGIYETNNVAVASPEKGAETWEDSYDCKFYLRTFNKLAERGAFFKNIETN 60

2y 214 YTRIFLYSGEPTLYGNETSVEGPTGNKTLGLAIKRFYPPKPHLPTEKFLSLIQIDAV 273
    61 YTRIFLYSGEPTLYGNETSVEGPTGNKTLGLAIKRFYPPKPHLPTEKFLSLIQIDAV 120

2y 274 IVHKQFLYFNFEYWFLEPMKFPFKITYEELPIRINKTISGL 316
    121 IVHKQFLYFNFEYWFLEPMKFPFKITYEELPIRINKTISGL 163

RESULT 6
JS-10-264-049-4034
; Sequence 4034, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264, 049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 4034
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
JS-10-264-049-4034

Query Match          26.8%; Score 472; DB 15; Length 89;
Best Local Similarity 98.9%; Pred. No. 2.9e-39;
Matches 88; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2y 228 GNETSVGPGNKTGLAIKRFYPPKPHLPTEKFLSLIQIDAVIVHKQFLYFNFEY 287
    1 GNETSVGPGNKTGLAIKRFYPPKPHLPTEKFLSLIQIDAVIVHKQFLYFNFEY 60

2y 288 WFLPMKFPFKITYEELPIRINKTISGL 316
    61 WFLPMKFPFKITYEELPIRINKTISGL 89

RESULT 7
JS-10-369-493-12714
; Sequence 12714, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
```

```
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12714
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-369-493-12714

Query Match          5.5%; Score 97.5; DB 15; Length 328;
Best Local Similarity 26.1%; Pred. No. 0.47;
Matches 41; Conservative 19; Mismatches 50; Indels 47; Gaps 7;

2y 138 TIVQVATISGNMENOQA-----KMWQDNNGIYETNNVAVASPEKGAET 182
    57 TIVSIGFQSNHTRQYAAVARKGLEVALVQEKWVMDKNGVYDAGNQLSRLMGADT 116

2y 183 WEDSYDCKFVLRTEFKLAFFGAE---PKNIETNYTRIFLYSGEPTLYGNETSVEGPTG 238
    117 RLD-----MGFGIEHKLTILKNE---EELIRAGGRKYI-----PAGASD 154

2y 239 NKTGLAIKRFYPPKPHLPTEKFLSLIQIDAVIV 275
    155 HPLGGLGFAR--WAFEVEMQEKELGVFFDTVIY 185

RESULT 8
US-10-408-765A-58
; Sequence 58, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale B.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 836
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-58

Query Match          5.4%; Score 95.5; DB 16; Length 836;
Best Local Similarity 20.5%; Pred. No. 2.7;
Matches 61; Conservative 39; Mismatches 125; Indels 73; Gaps 12;

2y 37 TQSPPIV-----MEGDDIEVFRLOAPVWEKYGDLG---HLKIMDAIGRSLNG 86
    272 TSAFSPAAGTVGMDNDEDDDFS-----KMFYIPQSSPPKLTIFKTLRVRSDER 324

2y 87 KNYTMEWELFOLGNCPTPHLRPEMDAPFWCNOGAACFEIGIDDVHKENG-TIVQVAT 145
    325 TQKVMWEELASGLLT-SLKDNV-----KATGVLDVYVKYMEHTGTLRVSSK 376

2y 146 SGMEFQMAKWKQDNNGIYETNNVAVASPEKGAETWEDSYDCKFVLRTEFKLAFFG 205
    377 LRRLTQDHAEWYQ---GAIREIDIDERFQAGAGTGTQV----- 415

2y 206 EFNKNIETN-YTRIFLYSGEPTLYGNETSVEGPTGNKTLGLAIK-----RF 249
    416 EWDKQNLXQBELILQEGASFGGLKDNVFDGLVVRTOGFHMKVKRLDISLIDFLNPRF 475

2y 250 YPEKPHLPTEKFLSLIQIDAVIVHKQFY-----LFYNEBYWFLPMKFPFK 298
```

b 476 QPGRKGLYREBELCTMF-IREVGTLSQVYSKVANGSEILFSYQPDVITLPELRK 532

RESULT 9
S-10-282-122A-58749

Sequence 58749, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 58749

LENGTH: 1238

TYPE: PRT

ORGANISM: Helicobacter pylori

S-10-282-122A-58749

Query Match 5.3%; Score 94; DB 12; Length 1238;

Best Local Similarity 23.6%; Pred. No. 6.4;

Matches 62; Conservative 28; Mismatches 77; Indels 96; Gaps 15;

80 FSTSTLTKNYTWEELFOLGCTFPHLRPEMDAPWCN-----QGAACFEFGIDVHWK 134

513 FQVITGGKNSALVF-----NATTP-----WANGSIPKSNSTVRFEGYGVWVG 555

135 ENGTLVOYAT-----ISGNMFMQMAKWKQDNETGIYETTNVVKASPEKGAFTWDSYDC 189

556 KGVYTTGFTADRVYITGNM-----TG-----NGAQTGGCA--T 588

190 SKFVLRTNKLAFAEFKNIEET--NYTRIPLYSGEPTYLGNETSVPFGPTNKTGLAI 246

589 INFVQATEINIA--GATFENKTKTSQNSYMTFMALGD-----SSGSAKINYSQ 634

247 KFFY-----YFKNHLPKEFELSLQIFDAVYIYHKOFLF-----YNFEYM--FLPMK 293

635 SDPYMTGGGYDFTGN-----GVFDSVNFKKAYYKFGQTSNSYFNPTNFTLAGN 683

Qy 294 FPF1-KITYEELPIPIRINKTSLG 315
Db 684 FKFGKTTIEKSVLSDSASTFDG 706

RESULT 10
US-10-282-122A-55615

Sequence 55615, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 55615

LENGTH: 478

TYPE: PRT

ORGANISM: Enterobacter cloacae

US-10-282-122A-55615

Query Match 5.3%; Score 93; DB 12; Length 478;

Best Local Similarity 22.1%; Pred. No. 2.2;

Matches 56; Conservative 28; Mismatches 77; Indels 92; Gaps 16;

92 EWE-----LFQGNCTFPHLRPEMDAPWCNQGAACFEGIDV--HW----- 133

211 DWYSNEBGRALDWMGVASLPRKL--DFQSPILVNE-----YGGEDSIYRHWLKAPWMDG 264

134 -----KENGTLVOYATIS-----GNMFMQMAKWKQDN-----N 161

265 WRLDVVNLGAGERRANNLOHVAGITRGAAKAQAEPAFVGEHFGDARQWLQNDADDAAMN 324

162 ETGIYETWNVKASEKG-----AET--WFDSDYDS--KFLVRFENKJ-----AEF 203

325 YRGFTFPMGLANTDISYDPNHDATCMAMENYRAGLSHQQLRNFQNDSDHTARF 384

204 ---GAEFKNIEYTRIIFYSGEP--TYLGNETSVEG--PTGNKTLGLAIRFYPPPK 255

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

run on: June 8, 2004, 13:59:38 ; Search time 22.8222 Seconds

(without alignments)
1351.883 Million cell updates/sec

title: US-10-010-050A-2_COPY_31_346

sequence: 1 RVSGIPSRHWPVYKRFDF.....IKIYVEIPIPIRKKTISGL 316

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 283366 seqs, 96191526 residues

total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95.5	5.4	359	2	A96816 F9K20.25 [imported
2	94	5.3	362	2	T33904 hypothetical prote
3	94	5.3	1238	2	A64596 hypothetical prote
4	93	5.3	561	2	T19604 hypothetical prote
5	93	5.3	3036	2	T18995 hypothetical prote
6	92	5.2	502	2	F70316 conserved hypotbet
7	91	5.2	291	2	T21702 hypothetical prote
8	91	5.2	508	2	D85056 probable aspartic
9	90	5.1	585	1	UC1486 neopululanase (EC
10	89	5.1	1353	2	T27404 hypothetical prote
11	88.5	5.0	465	2	T51094 acid phosphatase (
12	88.5	5.0	1462	1	D6HUAC DNA-directed DNA p
13	88	5.0	741	2	D69453 conserved hypotbet
14	88	5.0	412	2	S27263 synaptic vesicle p
15	88	5.0	3194	2	D71917 toxin-like outer m
16	87.5	5.0	755	2	T41912 structural phospho
17	87.5	5.0	1162	2	T37889 probable nuclear p
18	87	4.9	619	2	T27026 hypothetical prote
19	86.5	4.9	388	2	T49253 purple acid phosph
20	86	4.9	411	2	AF1352 probable phosphoe
21	86	4.9	1086	2	UC6079 chitin synthase (E
22	85	4.8	752	2	B59200 acid phosphatase (
23	84.5	4.8	464	1	B59200 acid phosphatase (
24	84.5	4.8	465	1	T51095 acid phosphatase (
25	84	4.8	247	2	H69493 phosphoesterase-re
26	83.5	4.7	441	2	F71425 hypothetical prote
27	83.5	4.7	984	2	T4496 cellulose 1,4-beta
28	83	4.7	228	2	H69399 hypothetical prote
29	83	4.7	552	2	E72283 alpha-galactosidas

30	83	4.7	620	2	T27008 hypothetical prote
31	83	4.7	653	2	F70383 organic solvent to
32	83	4.7	1086	2	T33893 hypothetical prote
33	83	4.7	4563	1	LEPHUB apolipoprotein B-1
34	82.5	4.7	403	2	AF3413 fosmidomycin resis
35	82.5	4.7	509	2	S49349 cytosin (EC 3.4.2
36	82.5	4.7	672	1	A41043 acetate-CoA ligase
37	82.5	4.7	704	2	B83914 beta-xylosidase (I
38	82	4.7	386	2	T02640 hypothetical prote
39	82	4.7	456	2	AH0842 probable glycopor
40	82	4.7	477	1	A36709 alpha-amylase (EC
41	82	4.7	553	2	T03858 hypothetical prote
42	82	4.7	720	2	G70320 aldehyde dehydroge
43	82	4.7	742	1	A43344 synaptic vesicle p
44	81.5	4.6	514	2	T39458 DNA binding protei
45	81.5	4.6	2167	2	G19444 hypothetical prote

ALIGNMENTS

RESULT 1

A96816
F9K20.25 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A96816
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzay, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.N.; Luo, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96816
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-359 <STO>
A:Cross-references: GB:AB051173; NID:G3834323; PIDN:ACG83039.1; GSPDB:GN00141
A:Genetics:
A:Gene: F9K20.25
A:Map position: 1

Query Match 5.4%, Score 95.5; DB 2; Length 359;
Best Local Similarity 21.5%; Pred. No. 1.8;
Matches 53; Conservative 39; Mismatches 97; Indels 57; Gaps 12;

QY	32	YTFPCTGSPIDVMEGDDIEVFLQAPVWEKYGDLGHEKIMHDAIGFRSTLGKNTYM	91
DB	80	YRMOPGTGDIIPRFNGRDFLFRFGKILF--VDSLSN-----N	116
QY	92	EWEYELPOLGNGTFRHLEPMOAPWCNOGACFF---EGIDVHWKENGTLVQAT----	144
DB	117	MWVSL-----SCMLHAAPVNAKYTFOLNKGLSTFTIPRYGI-SVNFLLKNGFLVDVSDKTR	171
QY	145	---ISGNMFNQMAKVQADNETGIYETWNVKASPEKGAETWFSYDSKFEVLRFTFNKLA	201
DB	172	GLILKLSISRGNQMLSDVAIFNTFFHW-----SHTGAKTW-DYFCGDKIVKEMNME	226
QY	202	EFGRGF-----KNIEYNTYRIFLYSGEFTYI-GNETSVFGPTGKNTLG--LAIRKFY	250
DB	227	AFKIALTTWSKMDINIDPSKTRVYGVSPVHNGE---WGRPKGKTLGETVAVQGPS	283
QY	251	YFPKPH 256	
DB	284	YFGKPN 289	

RESULT 2

Query Match 5.3%; Score 93; DB 2; Length 3036;
Best Local Similarity 23.2%; Pred. No.44;
Matches 76; Conservative 40; Mismatches 102; Indels 110; Gaps 2;

72 KIMDAIGRSTLTGKNYTMWEYELFQLCNCTP-----HLRPENDAFWNCOGAAACEF 125
|||::|:
1574 KILLDGYVAFANISSNNFTLN-----TAPKQFGWVEIR-GNEKULICERNPFEN 1622
|||::|:
126 ESIDDVHM--KENGT---LVQVATISGMNFQMAKKVKODNEGITYETWNVAASEPKGA 180
|||::|:
1623 WG----HMKIKDATSQYLROIADVPSYQYNYLEK-NREINRGDYVDWM-----PRSYA 1672
|||::|:
181 EDWPDSDYCSKRVLRFNFKLAEF-----GAEPKNI-ETNTY-----RIPLY 220
|||::|:
1673 LCVFSGSOKYEVIHPNRFPNELIDELVELYSAGKYSVFESTNNITNMWGTCNEAVIAQRFDF 1732
|||::|:
221 SG-----EPTVLGNETSIV-----FGETGN-----KTGLGA 245
|||::|:
1733 DDMNFTFAEMTPPYVSNDLSINFMMNPWRDGLAATYIEPCVHDLHGVEDKXULTLI 1792
|||::|:
246 IKRF-----YYPEKPHLPTEKEFL--SLQLQFDAYIVH---KQFYLYFNF-----EW 288
|||::|:
1793 TERWSEFPYVPBPFRERYRITRDVTIMPGATLYIEQWVEHVMPNVAILVLGNLARGEYW 1852
|||::|:
289 FLPMKPEPPFIKIT---YE-EIPLPIRNK 311
|||::|:
1853 -QPIRFKPINTEVMQYKEIPTEYRK 1879
|||::|:

Result 6
70316
onserved hypothetical protein ag_175 - Agulfex aeolicus
/Species: Agulfex aeolicus
/Date: 08-May-1998 #sequence_revision 08-May-1998 #ext_change 31-Mar-2000
/Accession: F70316
/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O

ature 392, 353-358, 1998
/Title: The complete genome of the hyperthermophilic bacterium Agulfex aeolicus.
/Reference number: A70300; MWID:98196666; PMID:9537320
/Accession: F70316
/Status: preliminary; nucleic acid sequence not shown; translation not shown
/Molecule type: DNA
/Residues: 1-502 <AOE>
/Cross-references: GB:AE000677; NID:g2982900; PIDN:AAC06534.1; PID:g2982916; GB:AE00065
/Experimental source: strain VFS
/Genetics:
/Gene: ag_175
/Superfamily: probable 60k inner membrane protein; stage III sporulation protein homolo
/293-486/Domain: stage III sporulation protein homology <SPOR>

Query Match 5.2%; Score 92; DB 2; Length 502;
Best Local Similarity 21.3%; Pred. No.54; Mismatches 114; Indels 68; Gaps 15;
Matches 63; Conservative 51;

40 PIPVMEGGDDIEVFRLQAFVMEFKYGDLLGHKIMDAIGRSTLTGKNYTMWEYELFOL 99
|||::|:
118 PLRIFTGNDLDQ-KLNFEYELKEGK--NSVALIHKLKVKILSYKNGAI----- 165
|||::|:
100 GNCTPPLRPE-MDAFWCNGQA---ACFESIDIVHKENGTLVQVATISGMNFQMA 154
|||::|:
167 -----HSVEGLKPEFWFVGSPDPDEAFYTHVGV-LKINEAVRLDV----- 203
|||::|:
155 KWVKQDNERTGIYETWNVAASPEKGAETWDSYDCSKFVLRTNKLAFFGAEPKNIETNY 214
|||::|:
210 ----DLKGINEFBECNIFPGGBESRYFFKGADQYKHVIYK---VKLGDKVSLST- 258
|||::|:
215 TRIFLYSGEPT-YLG-----NETSVFGPTGKTKLGLAK--REFYPKPHLPTEKEFL 263
|||::|:
259 ---FLYDGKTIYLGAKOVARLRRLGIVDTLMDWGLTKIIVKPLFLFYWIYEH--TSWV 313
|||::|:
264 ISLL-----QLPDANIYKQFLTFNFEYWFLEPMKPEPPFIKITIYEEPLPIRNKTL 313
|||::|:

Db 314 LSIVLFTVRIFFPLGYKSVSMQKIQE--LAPMEKIKQKXYKDDPVYMQGEMM 367

T21702
 RESULT 7
 Hypochemical protein F33E2.5 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T21702
 R:Lennard, N.
 submitted to the EMBL Data Library, January 1997
 A:Reference number: 219461
 A:Accession: T21702
 A:Status: preliminary; translated from GB/EMBL/DDDJ
 A:Molecule type: DNA
 A:Residues: 1-291 <WIL>
 A:Cross-references: EMBL:284574; PIDN:GAB06543.1; GSPDB:GN00019; CESP:F33E2.5
 A:Experimental source: clone F33E2
 C:Genetics:
 A:Gene: CESP:F33E2.5
 A:Map position: 1
 A:Introns: 252/1

Query Match 5.2% Score 91; DB 2; Length 291;
 Best Local Similarity 17.9%; Pred. No. 3.3; Matches 57; Conservative 47; Mismatches 92; Indels 122; Gaps 13;

QY	27	YQAKYTECP----	TGSPIPWEGDDIEVFRL-QAPWEFKYGDLLGDKIMDAIGFR	81
		: :	: :	: :
Db	25	YPEARVPTPKIKKTA	DQKLEGSKDIRIAFTLSQEPIDAKCTEMTKNVLLQ----	FQ 80
QY	82	STLGKNT-----	TMWYELFQLGNCSTPHLRPEMDAPWCNOGA	121
		:	:	:
Db	81	DTDIPANYIAVSMALP	NFKIGQNCHEFDPLDWYKNIK-GYFTYKKITEQVQAAYFNDEAG	139
QY	122	ACFEFGIDDMHKEGTL	VQVATISGNMFNQMAKWK--ODNETGIYETMNVAASPERG	179
		:	:	:
Db	140	NQFLAGV-----	CYATPAG-----KWAQKIQTEQTLRFEF--QTAEP--	176
QY	180	AETWPDSDSCFVLRT	ENKLAEGAFENKIETNYRIFLYSGEPTYLGNETSVEFGPTGN	239
		: :	: :	: :
Db	177	----FPLNSKSRFDIK	APFKFARYG--RYCEEGGRVLWVGK-----	214
QY	240	KTLGLAIRFYYPFKPH	LPTKEFLSLQIPDAIVYHKQFLFY-----NF	285
		:	:	:
Db	215	-----	DGQIVHLNHYVCMQWTRGCACAKDHF	243
QY	286	EYWFPLMKFPFIKIT	YEE 303	
		: :	: :	
Db	244	EYMWIAPEEPVYQSK	EE 261	

RESULT 8
 DB5056
 Probable aspartic proteinase [imported] - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
 C:Accession: DB5056
 R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
 Nature 402, 769-777, 1999
 A>Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.
 A:Reference number: A85001; MUID:20083488; PMID:10617198
 A:Accession: DB5056
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-508 <STO>
 A:Cross-references: GB:NC_001268; NID:g7267203; PIDN:CA877914.1; GSPDB:GN00140
 C:Genetics:
 A:Gene: AT4G04460
 A:Map position: 4
 C:Superfamily: oryzasin; saposin repeat homology

Query Match 5.2%; Score 91; DB 2; Length 508;

```

Best local similarity 23.8%, Pred. No. 6.7;
Matches 36; Conservative 17; Mismatches 48; Indels 50; Gaps 7;

yb      27 YCOAKY-----TCPTGSPIPVMEG-----DDLEVFRLQAPVEF-----62
           ||::||::||::||::||::||::||::||::||::||::||::||::||
yb     125 YHSHKTKKSSSSSTRKKGKPASIVYGALSGIFSNDDVVGIVVKEDFLFATSEPI 184
           ||::||::||::||::||::||::||::||::||::||::||::||
yb      63 -----KYGDLLGHKKIMDAIGFRSTLTGNRYTWEMWELETOLGNCTPHLRPEMDAPENC 117
           ||::||::||::||::||::||::||::||::||::||::||::||
yb     185 TPLLAKFPGLTG-----LGFKEISVG-NSIFWVMINVEKG-----LVKKEPLPSFWL 229
           ||::||::||::||::||::||::||::||::||::||::||::||
yb     118 N-----OGAACFFEGIDVDVHMKNKGLTVQV 142
           ||::||::||::||::||::||::||::||::||::||::||::||
yb     230 NNENPKDPEGGELIVFGGVDPKHFKKEHTFFEVV 260

```

```

RESULT 9
JC1486
neopullulanase (EC 3.2.1.135) - Thermoactinomyces vulgaris
4/Alternate names: alpha-amylase II
4/Species: Thermoactinomyces vulgaris
4/Date: 31-Dec-1993 #sequence_revision 18-Aug-1995 #text_change 16-Jun-2000
2/Accession: JC1486
3/Tonowaka, T.; Ohtsuka, M.; Mogi, S.; Sakai, H.; Ohta, T.; Sakano, Y.
3/Source: Biotechnol. Biochem. 57, 395-401, 1993
4/Title: A neopullulanase-type alpha-amylase gene from Thermoactinomyces vulgaris R-47.
4/Reference number: JC1486; MIMD:93222535; PMID:7763540
4/Accession: JC1486
4/Molecule type: DNA
4/Residues: 1-585 <TON>
4/Cross-references: GB:013178; NID:g391625; PTDN:BA02473.1; PID:g398125
3/Function:
4/Description: Hydrolysis of alpha-(1->4)-glucosidic linkages of pullulan to produce pan
4/Pathway: pullulan degradation
1/Note: also has alpha-amylase activity
2/Superfamily: neopullulanase; alpha-amylase core homology
4/Keywords: glycosidase; hydrolase; polysaccharide degradation
4/3293-424/Domain: alpha-amylase core homology <AM1>
4/3293-354/421/Active site: Glu, Gnu, Asp #status Predicted

```

Query Match	5.1%;	Score 90;	DB 1;	Length 585;
Best Local Similarity	21.5%;	Pred. No. 9.7;		
Matches	53;	Conservative	30;	Mismatches 68; Indels 96; Gaps 15

```

2y      71 LKIMEDAD-----GERSTLT--GKNTMTME-----YELFOLGNCNTE 104
      234 IKILIDAFENHAGDFFAFNRDLVKGKQSKRYDMFFIEDPVSUKTSRTNYETFAQVAPAM 253
2y      105 PHLR---PEMD-----APFWCNGAACCFEG-----IDDVHKENGTLVQV---A 143
      294 PLKLRTEPNEVEKEYLEDVAFPMEEGG---IDGMRILDVANEVDHAFMRPRRLVKSINPDA 349
2b      144 TTSGMRFQMAKMYKQDNMTGI-----YYEPMNYKAS-----PEKG 179
      350 LTVGEIMHDASGMILMDGQFDSVMNYYLFRSEYIRFPATGTHARFPABELTRARMLTPEDA 409
2b      180 AE-TW--PDSYDCKSFVYLRTEKNKLAEFGAEKPNLEIYTRI-----FLYSGED-TYIGNE 230
      410 AAGLNNLLDSDHTERFLTSCGNEAKF-----RLAVLPQMTYLGPIIYYGDE 457
2y      231 TSVFGPT 237
2b      458 IGMAGAT 464

```

RESULT 10
T27404
hypothetical protein Y75B8A.22 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27404
R:Parlow, K.
submitted to the EMBL Data Library, November 1998

A:Reference number: Z20361
A:Accession: T27404
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1353 <MIL>
A:Cross-references: EMBL:AL03514; NID:e1343251; PID:CAA22106.1; CESP:Y75B8A.22
A:Experimental source: clone Y75B8A
C:Genetics:
A:Gene: CESP:Y75B8A.22
;Introns: 28/1, 79/2; 634/3, 996/1, 1134/1; 1296/3

```

Query Match 5.1k; Score 89; DB 2; Length 1353;
Best Local Similarity 19.1k; Pred. No. 34;
Matches 73; Conservative 51; Mismatches 124; Indels 134; Gaps 18;

QY      1 RVSGIP-----SRHHMPVYKR--FDRPKDPYCO-----AK 31
Db      607 RSRGFPAAVGLYHNASRALIMPESFRGGLTDQDSGEGDQQLQETQLKADMKVAKDLKK 666

QY      32 YTFPGTSSPIRVNEGDDIEVRLQAVWEFKYDGLGLKIMDAIGFSTLGGKYTM 91
Db      667 AESKCTDEDPAVKKYKRMATAQS-LWE-QSTDLARI-LSHELPESEST-SPVVMQL 722

QY      92 EWEYLFOLGNCTPEHLREPMDAQ-----WCNOGAACF 124
Db      723 D-----IRPDVOQKRAMALIQALREARDLPAAVGLYHNSRKLMPEDEALFG 768

QY      125 PEGIDVAHMKENGTLVQATISGM--PNOAKMYK-----QDNETGIYETNNYKAS 175
Db      769 APGI-----GVEEIEALIKALLEADLHEVAEMKVAEDRADPDPEEDPAEYDSEOE 820

QY      176 PEKAETM-----PDSYDCKFVARTENKLAEGAEFKNLEETNTATRLFLVSGEPTYL 227
Db      821 EEELVPAMKABEIDFQDSYVC-----KSNVDLKMVYELN--DFS 861

QY      228 GNETSVGPTGNKTLGLAIRFYYPKPHLPYKEFLSLIQIPDAVTVH-----KQ 278
Db      862 KNSTEL-----NQALVLMRLHRIAPDLK-LPIKLQVSLFQVFSKVENHEFTHLSKDLRKS 914

QY      279 FYLFYNEFYMFLPKKFPPIKIT 300
Db      915 SRLLEYQFGHHLTKKEFSKPT 936

Db

```

RESULT 11
T51094
acid phosphatase (EC 3.1.3.2) purple 1, precursor [similarity] - sweet potato
C:Species: Ipomoea batatas (sweet potato)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 22-Oct-2001
C:Accession: T51094
R:Dumas, A.; Bicken, C.; Spener, F.; Krebs, B.
Biochim. Biophys. Acta 1434, 202-209, 1999
A:Title: Cloning and comparative protein modeling of two purple acid phosphatase isozymes
A:Reference number: Z25292; MUID:20028255; PMID:10556574
A:Accession: T51094
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-465 <DUR>
A:Cross-references: EMBL:AJ006224; PIDN:CAA06921.1
C:Genetics:
A:Gene: pap
C:Superfamily: kidney bean purple acid phosphatase; phosphoesterase core homology
C:Keywords: phosphoric monoester hydrolase

```

Query March 5.0%: Score 88.5; DB 2; Length 465;
Best Local Similarity 21.4%: 1 Pred. No. 9.7;
Matches 51; Conservative 31; Mismatches 77; Indels 79; Gaps 13;

QY      131 VMKENGIVQVATISGM-----FNQAAKVKRQ-----DNELGIYET-----W 170
      1: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      87 VMSENGQHKVLA--KGIKITYTFNTSGYIHHCITRNLDYNTKYYEYVGIGNTSRFSW 144
      1: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      171 NVKASPEKGAETWDSVDCKSKFYIKRTNKLAEPFAEKXNIE--TNYR-----IFLXG 222

```

```

b      145 -FTTPEVGPDP-----VPTFTGLIGDLSGFSDFSNRFLTHYERNPIKQAVLFEVG 192
Y      223 EPTVYGN-----ETSVFGR-----TGNKTTLGA-----IKRF 249
b      193 DLSYADNYNHNDVNMOTMGRFVNERSTAYQPMWTAGNEIDLPAPETIGTKPKPKPTKXY 252
Y      250 YYPFKFHLPTKEFLLSLQIFDAVIVHKQF--YLEFNYFEWFLPMKFPFIKITYEIRP 305
b      253 HVPYKASGSTETFTFWYSIKRASAVIIVLSSYSAYGKKTTPQYKNLEELTF--KYNRTETP 308

```

RESULT 12

CHUAC

NA-directed DNA polymerase (EC 2.7.7.7) alpha catalytic chain - human

```

Species: Homo sapiens (man)
Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 27-Oct-2003
Accession: S00257; A04040; I57513; S12665
Wong, S.W.; Wahl, A.F.; Yun, P.M.; Pearson, B.E.; Arai, K.I.; Korn, D.; Hun
MBO J. 7, 37-47, 1988
Title: Human DNA polymerase alpha gene expression is cell proliferation dependent and
Reference number: S00257; MUID:88196090; PMID:3359994
Accession: S00257
Molecule type: mRNA
Residues: 1-1462 <WON>
Cross-references: EMBL:X06745; NID:g35567; PIDN:CAA29920.1; PID:g35568
Accession: A10440
Molecule type: protein
Residues: 438-449;495-502; 'G', 838-848;1090-1105;1201-1216;1397-1407;1444-1453 <MON2>
Pearson, B.E.; Nashener, H.P.; Wang, T.S.
Ol. Cell. Biol. 11, 2081-2095, 1991
Title: Human DNA polymerase alpha gene: sequences controlling expression in cycling an
Reference number: I57513; MUID:91172197; PMID:2005899
Accession: I57513
Status: translation not shown; translated from GB/EMBL/DBJ
Molecule type: DNA
Residues: 1-8 <RES>
Cross-references: GB:M64481; NID:9181617; PIDN:AAA52318.1; PID:g181618
Hsi, K.L.; Copeland, W.C.; Wang, T.S.F.
Nucleic Acids Res. 18, 6231-6237, 1990
Title: Human DNA polymerase alpha catalytic polypeptide binds ConA and RCA and contain
Reference number: S12665; MUID:91057099; PMID:2243771
Accession: S12665
Molecule type: protein
Residues: 19-37; 'C',1406-1425, 'C' <HSI>
Genetics:
Gene: GDB:POLA
Cross-references: GDB:120304; OMIM:312040
Map position: Xp22.3-Xp21.1
Superfamily: DNA polymerase
Keywords: DNA binding; DNA replication; nucleotidyltransferase; zinc finger
654-691/Region: zinc finger CHCC motif
1249-1374/Region: zinc fingers

```

```

Query Match      5.0%; Score 88.5; DB 1; Length 1462;
Best Local Similarity 20.3%; Pred. No. 42;
Matches 71; Conservative 56; Mismatches 115; Indels 107; Gaps 20;

```

```

Y      16 KRDFRPKPDYPCQAKYTFPCPTG-----SPIVWEGD---DIEIV-----FRL 55
b      183 KRKRSIGASPNPFSVTATAVPSGKASPVSRKEPPLTVPPLKRAFPADDDQVESTEEQ 242
Y      56 QAPVVEFYGDLLGLKIMHDAIGRSLTIGKNYTMWEYELFQJGNCFPHLRPMADPF 116
b      243 ESGAVEFEDGDP-----DEMEVEEV-----DLEP-MAKA 272
Y      116 WCNQACAFCEFGIDVHWKEN---GLIVQVATISGNMFMQAKM-VKQDNETGIYETWN 171
b      273 WDKES-----EPAEYVKQKADSGKGTV-----SYLGSFLPDVSCMDIDEGSSSESVQEVQ 323
Y      172 VKASP---EKGAET-----WFDSDYD---CSKVFYLFKFKIAEFGAE-----FFNI 210
b      324 VDSSTPLVKGADEQVHFHFWLDAVEQYQNPQGVVFLFGKWTLSAEETHVSCVMVXNI 383

```

```

QY      211 ENNYTRIFLYSGEPYTLGNETSVFPGTGNKT-LGLAIRFYPPFKPHLPTKEFLLSLDOI 269
Db      384 ERT-----LY-----FLPREMKIDINTGKTGTPTSMKDVYEEPEKATKYIKM----- 429
QY      270 FDAVIVHKQFYLYNFYEWFLPMKFPFIKITYE-----EILPIRNKTS 314
Db      430 FRSKRVENK-----YAFELPDVPEKSEYLEVKKYSAEMPQLPOLKKEITFS 414

```

RESULT 13

D69459

Conserved hypothetical protein AF1677 - Archaeoglobus fulgidus

```

C/Species: Archaeoglobus fulgidus
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 17-Mar-2000
Accession: D69459
R.Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A/Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Arriach, P.; Kaine, B.P.; Sykes, S.B.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A/Reference number: A69250; MUID:98049343; PMID:9389475
Accession: D69459
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
Molecule type: DNA
Residues: 1-411 <KLE>
Cross-references: GB:AE000987; GB:AE000782; NID:g2689310; PIDN:AAB89568.1; PID:g2648873
Superfamily: hypothetical protein H10701

```

```

Query Match      5.0%; Score 88; DB 2; Length 411;
Best Local Similarity 20.3%; Pred. No. 9.2;
Matches 55; Conservative 45; Mismatches 97; Indels 74; Gaps 13;

```

```

QY      44 MEGDDIEVFRLQAPVWEFYGDLLGLKIMHDAIG-----FSTLTGKNYTMWEYEL 96
Db      38 LSDEGDFLIIRVEKKNM-----DTLNFARVLSNMLGISQKRSFAGTCDKRALTYQYFET 92
QY      97 FQLNGCTFPHLRPEKADPFQCNQACAFCEFGIDVHWKE-----NGTLVQVATISGN 148
Db      93 YGVKK-----FEIERYNLDAKIEVIGYARRAQLQDGLLGN 128
QY      149 MFMQAKMVKQDNENGTIYETWNVAKSPKQAEWTF--DSVDCSFVLRFNKLAFFGAE 206
Db      129 FF-RIRVYGCDDGE--TFQETRN--ELMEKGTPNVFGIQRFGSITFTHEVGKLT----- 178
QY      207 FKNIEIWTYTRIF-LYSGEPYTLGNE-----TSVFGPTGNKTLGLAIRFYPPFKPHLPTK 260
Db      179 ---LQNNVEAFWYVAKPFEGNEBVKIKREILIMFPDACLGLRELPKYLRVENLLOK 235
QY      261 -----EFLLSLLQIFDAVIVHK-QFYLF 282
Db      236 LREKSEEEALLSLPKNLKMMFVHAYQSYIF 266

```

RESULT 14

S27263

synaptic vesicle protein SV2 - bovine

N/Alternate names: transporter-like protein p87

C/Species: Bos primigenius taurus (cattle)

C/Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 18-Aug-2000

Accession: S27263

R.Gingrich, J.A.; Andersen, P.H.; Tiberi, M.; El Mestilawy, S.; Jorgensen, P.N.; Fremeanu

FBS Lett. 312, 115-122, 1992

A/Title: Identification, characterization, and molecular cloning of a novel transporter-1

A/Reference number: S27263; MUID:99050176; PMID:14426240

Accession: S27263

Molecule type: mRNA

Residues: 1-742 <GIN>

Cross-references: GB:S47919; NID:g259173; PIDN:AAB24028.1; PID:g259174

Superfamily: synaptic vesicle protein SV2

Query Match 5.0%; Score 88; DB 2; Length 742;
Best Local Similarity 22.1%; Pred. No. 19;
Matches 57; Conservative 36; Mismatches 97; Indels 68; Gaps 14;

```

|y 86 GRKYTWEMWYLLFQIGNCTFPH-----LRPEMDAP-FMCMQSG---AACFPEGIDVDVHWK 134
|b 325 GSAYQFHSWVFWL-VCAFPVSFAIGALTQPSRPFLENGKHDEAMWYLKQVHDTNMR 383
|y 135 ENGTLVQVATISG-NMFMQAKWVKODNETGIYETWNVKASPEKGAETWPDSDCKFY 193
|b 384 AKGHPERVSVTIKTHQDELEIQTSDGAMVQRMGVHAL-SLGGQVWGNFLSC----- 438
|y 194 LRTFNKLAIEFGAEFKNIETN---YTRIFLYSGEPTYLGN-----ETSVF-- 234
|b 439 -----FGPEYRRLTLMWGMWFTMSGYGLTWPFPMIRHLQAVDYAARTKVPFG 489
|y 235 ---GPTNKTGLAIKRPYPPKP---HLPTKEFLSLQ--IFDAVIYHKQFY--- 283
|b 490 ERVEHYTFNFTLENOIHRGGQYFNDRKFIQLRLKSVSEFEDSLFEQCYFEDVTSNTFFRNC 543
|y 281 ----LFTN---PEYWEPL 290
|b 550 TFINVVFYNTDLPEYKGV 567
```

RESULT 15
71917
:okin-like outer membrane protein jhp0556 - Helicobacter pylori (strain J99)
:Species: Helicobacter pylori
:Variety: strain J99
:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
:Accession: D71917
:Alm, R.A.; Jhing, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
:Ives, C.; Gibson, R.; Metberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
:ature, 397, 176-180, 1999
:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
:Reference number: A71800; PMID:99120557; PMID:9923682
:Accession: D71917
:Status: preliminary
:Molecule type: DNA
:Residues: 1-3194 <ARN>
:Cross-references: GB:AE001468; GB:AE001439; NID:94155100; PIDN:AAD06134.1; PID:9415510
:Experimental source: strain J99
:Genetics:
:Gene: jhp0556

Query Match 5.0%; Score 88; DB 2; Length 3194;
Best Local Similarity 23.2%; Pred. No. 1.2e+02;
Matches 58; Conservative 29; Mismatches 69; Indels 94; Gaps 15;

```

|y 80 FRETLLGKNVYTWEMWYLLFQIGNCTFPHLRPEMDAPFMCN-----QGAQCFEGIDVDVHWK 134
|b 513 FAQTYGAKSALVF-----NATTP-----MANGAIPKSNSTYRFGYBGVNWG 555
|y 135 ENGTLVQVAT---ISGNFMQAKWVKODNETGIYETWNVKASPEKGAETWPDSDYDC 189
|b 556 KTGVIINGTFTADRVYITGNMS-----GNGAQGTGGA--T 588
|y 190 SKVLTLPFNKLAIEFGAEFKNIETNTRIFLYSGEPTY--LGNETSVPFGPTGNKTGLAIK 247
|b 589 LNFVGATEINIA--GATFNKLTKTSONSYW---TFMALGN-----GSGSGKINVSOS 635
|y 248 RFY-----YFEPKPHLPTKRELLSLQIFDAVIYHKQFYF-----YNFEYV-FLPMKF 294
|b 636 DFYDMTDGSGIDFTGN-----GVFDSVNFKAIFYKFGAENSYNFKNTNPLAGNF 684
|y 295 PFI-KITYEE 303
|b 685 KFOGKTIIEK 694
```

Search completed: June 8, 2004, 14:08:10
Job time : 23.8222 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

* protein - protein search, using sw model

on: June 8, 2004, 13:56:38 ; Search time 13.1667 seconds

(without alignments)
1249.684 Million cell updates/sec

title: US-10-010-050a-2_COPY_31_346
effect score: 1758
sequence: 1 RVSGIPSRHWVPYKRFDF.....IKITYEIPLPTRNKLTSL 316

scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

sarched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

imum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1758	100.0	407	1	CLN5_HUMAN
2	92	5.2	502	1	OXNA_AQUAE
3	90	5.1	585	1	NEP2_TREUV
4	89.5	5.0	427	1	ARP3_ACA
5	88.5	5.0	1462	1	DPOA_HUMAN
6	88	5.0	411	1	TRUD_ARCTU
7	87.5	5.0	755	1	P100_HSVJ
8	87	4.9	1184	1	CHSD_EMENT
9	85	4.8	254	1	SSA2_MOUSE
10	83.5	4.7	2167	1	BPH1_YEAST
11	83	4.7	228	1	YC01_ARCTU
12	83	4.7	889	1	CI22_HUMAN
13	83	4.7	999	1	HGPB_HAETN
14	83	4.7	1028	1	FDXG_HAETN
15	83	4.7	4563	1	APB_HUMAN
16	82.5	4.7	564	1	TM16_HUMAN
17	82.5	4.7	672	1	ACGA_METSO
18	82	4.7	2492	1	ATRX_HUMAN
19	82	4.7	2492	1	ATRX_PANTR
20	81.5	4.6	514	1	CHB2_SCHPO
21	81.5	4.6	606	1	VE1_HPV7
22	81	4.6	644	1	URCM_RAT
23	81	4.6	1451	1	DPOA_RAT
24	80.5	4.6	282	1	HCHA_ECO57
25	80.5	4.6	544	1	OXNA_BORBU
26	80.5	4.6	755	1	POLG_DENIT
27	80.5	4.6	738	1	SEC6_DROME
28	80.5	4.6	835	1	VPS_ROTSL
29	80	4.6	509	1	APPI_ORYSA
30	80	4.6	2492	1	ATRX_PONPR
31	79.5	4.5	511	1	SYFA_METHH
32	79.5	4.5	603	1	VE1_HPV8
33	79	4.5	508	1	ASPR_HORVU

34	79	4.5	682	1	TDR5_HUMAN	Q8nat2 homo sapien
35	79	4.5	928	1	ODO1_RICCN	Q92j42 rickettsia
36	79	4.5	2292	1	POLG_EMCVB	P17593 e genome po
37	79	4.5	2292	1	POLG_EMCVB	P17594 e genome po
38	78.5	4.5	218	1	VIXS_BPK3	P10393 bacteriopia
39	78.5	4.5	282	1	HCHA_ECOL6	P59331 escherichia
40	78.5	4.5	282	1	HCHA_ECOL1	P31658 escherichia
41	78.5	4.5	643	1	HCHA_SHIFL	P59332 shigella fl
42	78.5	4.5	643	1	RHNI_MOUSE	O61085 mus musculu
43	78.5	4.5	660	1	AMY_BACSU	P00691 bacillus su
44	78.5	4.5	1764	1	YK84_YEAST	P34241 saccharomyc
45	78.5	4.5	2358	1	WOKD_SCHPO	Q9y719 schizosach

ALIGNMENTS

```

RESULT 1
CLN5_HUMAN
ID CLN5_HUMAN STANDARD; PRT; 407 AA.
AC 075503;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ceroid-lipofuscinosis neuronal protein 5 (CLN5 protein).
GN CLN5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A., VARIANT VLINCL ASN-279, AND VARIANT ARG-368.
RC TISSUE=Fetal brain;
RX MEDLINE=98324783; PubMed=9662406;
RA Savukoski M., Kloockars T., Holmberg V., Santavuori P., Lander E.S.,
RA Peltonen L.;
RT "CLN5, a novel gene encoding a putative transmembrane protein mutated
RT in Finnish variant late infantile neuronal ceroid lipofuscinosis.";
RL Nat. Genet. 19:286-288(1998).
RN (2)
RP SUBCELLULAR LOCATION, AND GLYCOSYLATION.
RX MEDLINE=21968572; PubMed=11971870;
RA Isoomppi J., Vesa J., Jalanko A., Peltonen L.;
RT "Lysosomal localization of the neuronal ceroid lipofuscinosis CLN5
RT protein.";
RL Hum. Mol. Genet. 11:885-891(2002).
CC -!- FUNCTION: Not known.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- PTM: Glycosylated.
CC -!- DISEASE: Defects in CLN5 are the cause of Finnish variant late-
CC infantile neuronal ceroid lipofuscinosis (VLINCL) [MTM:256731];
CC also known as ceroid lipofuscinosis neuronal 5 (CLN5). VLINCL is a
CC fatal childhood neurodegenerative disease characterized by
CC progressive visual and mental decline, motor disturbance, epilepsy
CC and behavioral changes. The first symptom is motor clumsiness,
CC followed by progressive visual failure, mental and motor
CC deterioration and later by myoclonia and seizures.
CC -!- DATABASE: NAME=NCL CLN5;
CC NOTE=Neural Ceroid Lipofuscinoses mutation db;
CC WWW="http://www.ucl.ac.uk/ncl/CLN5.html".
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF068227; AAC27614.1; -
CC Genew; HGNC:2076; CLN5.
CC MIM; 608102; -.

```

```

MIM; 256731; -.
GO; GO:0016021; C: integral to membrane; TAS.
GO; GO:0008151; P: cell growth and/or maintenance; TAS.
M Transmembrane; Lysosome; Glycoprotein; Neuronal ceroid lipofuscinosis;
M Disease mutation; Polymorphism; Epilepsy.
M TRANSMEM 75 91
M CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
M VARIANT 279 279 D -> N (in VLINC).
M VARIANT 279 279 /FtId=VAR_005137.
M VARIANT 368 368 K -> R (in dBSN:1800209).
M VARIANT 368 368 /FtId=VAR_005138.
M SEQUENCE 407 AA; 46339 MW; 449702DIDC9HFE84 CRC64;

Query Match 100.0%; Score 1758; DB 1; Length 407;
Best Local Similarity 100.0%; Pred. No. 1.3e-143;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 RVSGIPSRHMPVYKRFDPKDPYCAKYTCFSPDPVMEGDDIEVRLQAPW 60
Y 92 RVSGIPSRHMPVYKRFDPKDPYCAKYTCFSPDPVMEGDDIEVRLQAPW 151
Y 61 EFYKGLGLKLMHDAIGRSTLTGKNTMYEYELFOLGNCFFPHLRPMDEAFMONG 120
Y 152 EFYKGLGLKLMHDAIGRSTLTGKNTMYEYELFOLGNCFFPHLRPMDEAFMONG 211
Y 121 AACFFEGIDVHMKENGLVQVATISGNMNOAKAWKODNETGIYETMNVASPEKA 180
Y 212 AACFFEGIDVHMKENGLVQVATISGNMNOAKAWKODNETGIYETMNVASPEKA 271
Y 181 ETWEDSYDCSKFVLRFNKLAEFGAEFNKLEIYTRIFLYSGEPTYLGNESYCEPTGN 240
Y 272 ETWEDSYDCSKFVLRFNKLAEFGAEFNKLEIYTRIFLYSGEPTYLGNESYCEPTGN 331
Y 241 TLGLAKRFYPRKPHIPTYEELSLIQTDAVIVHKQFLPNEFYWFLPMKFPPIKTT 300
Y 332 TLGLAKRFYPRKPHIPTYEELSLIQTDAVIVHKQFLPNEFYWFLPMKFPPIKTT 391
Y 301 YEELPIRKTKLSGL 316
Y 392 YEELPIRKTKLSGL 407

RESULT 2
XAA_AQUAE STANDARD; PRT; 502 AA.
ID 06561;
Y 15-MAR-2004 (Rel. 43, Created)
Y 15-MAR-2004 (Rel. 43, Last sequence update)
Y 15-MAR-2004 (Rel. 43, Last annotation update)
Y Inner membrane protein oxal.
Y OXAL OR AQ_175.
Y Aquifex aeolicus.
Y Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
Y NCBI_TaxID=63363;
Y [1]
Y SEQUENCE FROM N.A.
Y STRAIN=VF5;
Y MEDLINE=98196666; PubMed=9537320;
Y Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
Y Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,
Y Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
Y "The complete genome of the hyperthermophilic bacterium Aquifex
Y aeolicus";
Y Nature 382:353-358(1998).
Y - FUNCTION: Required for the insertion of integral membrane proteins
Y into the membrane. Probably plays an essential role in the

```

```

CC integration of proteins of the respiratory chain complexes.
CC Involved in integration of membrane proteins that insert
CC independently and independently of the Sec translocase complex (By
CC similarity).
CC SUBUNIT: Specifically interacts with transmembrane segments of
CC nascent integral membrane proteins during membrane integration (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -1- SIMILARITY: Belongs to the OXA1/oxa family. Subfamily 1.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB00677; AAC06534.1; -.
CC PIR; F70316; F70316.
CC HAMAP; MF_01810; -.
CC DR InterPro; IPR01708; 60kDa_innermem.
CC DR Pfam; PF02096; 60KD_IMP; 1.
CC KM Transmembrane; Inner membrane; Complete proteome.
CC FT TRANSMEM 12 34 POTENTIAL.
CC FT TRANSMEM 308 330 POTENTIAL.
CC FT TRANSMEM 378 400 POTENTIAL.
CC FT TRANSMEM 447 469 POTENTIAL.
CC SQ SEQUENCE 502 AA; 58224 MW; 6C24B8405F2570AD CRC64;

Query Match 5.2%; Score 92; DB 1; Length 502;
Best Local Similarity 21.3%; Pred. No. 2.2; Indels 68; Gaps 15;
Matches 63; Conservative 51; Mismatches 114; Indels 68; Gaps 15;

Q 40 P1PMEGDDIEVRLQAPWMEFYKGLGLKLMHDAIGRSTLTGKNTMYEYELFOL 99
Q 118 PLKIFGNGLDQ-KLNFGEYELKEK--NSVELHKEKLVKKIISYKGA1----- 166
Q 100 GNCFFPHLRPE-MDAPFWCNGA----ACFFEGIDVHMKENGLVQVATISGNMNOA 154
Q 167 -----HLSVEGLKPFVFWVGSPPDEAFYTHGVV-LKINGEVVRLDV----- 209
Q 155 KMWKODNETGIYETMNVASPEKGAETWEDSYDCSKFVLRFNKLAEFGAEFNKLEIY 214
Q 210 -----DOLKINFEBSNIEFGGBESRYFFKGAQDYKAIYK---VKLGDFVSLST-- 258
Q 215 TRIFLYSGEPT-YLG-----NETSVFGPTGNKTLGLAIK--RFYFPKPHLPTKEPL 263
Q 259 ----FLYDGEXTIYLGAKDYARLRRLGLVDTLDMGTLKIIVKPLFLVYIEH--TGSWV 313
Q 264 LSLVLTFTVIRIFLPLGYKSVSMOKLOE--LAPMKIKKKYKDDPVKMOEEM 367
Q 314 LSLVLTFTVIRIFLPLGYKSVSMOKLOE--LAPMKIKKKYKDDPVKMOEEM 367

RESULT 3
NEP2_THEVU STANDARD; PRT; 585 AA.
ID 008751;
Y 01-NOV-1995 (Rel. 32, Created)
Y 01-NOV-1995 (Rel. 32, Last sequence update)
Y 15-MAR-2004 (Rel. 43, Last annotation update)
Y Neopullulanase 2 (EC 3.2.1.135) (alpha-amylase II) (TVA II).
Y TVALI.
Y Thermoactinomyces vulgaris.
Y Bacteria; Firmicutes; Bacillales; Thermoactinomycetaceae;
Y Thermoactinomycetes.
Y NCBI_TaxID=2026;
Y [1]
Y SEQUENCE FROM N.A., AND CHARACTERIZATION.
Y STRAIN=R-47;
Y MEDLINE=93222535; PubMed=7763540;

```

Tonozuka T., Ohtsuka M., Mogi S.-I., Sakai H., Ohta T., Sakano Y.,
"A neopululanase-type alpha-amylase gene from *Thermoactinomyces*
vulgaris R-47.";
Biosci. Biotechnol. Biochem. 57:395-401(1993).
[2]
X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
STRAIN=R-47;
MEDLINE=99241045; PubMed=10222200;
Kamitori S., Kondo S., Okuyama K., Yokota T., Shimura Y., Tonozuka T.,
Sakano Y.,
"Crystal structure of *Thermoactinomyces vulgaris* R-47 alpha-amylase II
(TVAII) hydrolyzing cyclodextrins and pullulan at 2.6-A resolution.";
J. Mol. Biol. 287:907-921(1999).
[3]
X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
MEDLINE=22047855; PubMed=12051850;
Kamitori S., Abe A., Ohtsuka A., Kaji A., Tonozuka T., Sakano Y.,
"Crystal structures and structural comparison of *Thermoactinomyces*
vulgaris R-47 alpha-amylase I (TVAI) at 1.6 A resolution and
alpha-amylase 2 (TVAII) at 2.3 A resolution.";
J. Mol. Biol. 318:443-453(2002).
-!- FUNCTION: Hydrolyzes pullulan efficiently but only a small amount
of starch. Endohydrolysis of 1,4-alpha-glucosidic linkages in
pullulan to form panose. Cleaves also (1-6)-alpha-glucosidic
linkages to form maltotriose.
-!- CATALYTIC ACTIVITY: Hydrolysis of pullulan to panose (6-alpha-D-
glucosylmallose).
-!- COFACTOR: Birds 1 calcium ion per subunit.
-!- SUBUNIT: Monomer.
-!- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D13178; BA002473.1; -.
PIR; JCI486; JCI486.
PDB; 1BVZ; 02-MAR-99.
PDB; 1G1Y; 14-MAR-01.
PDB; 1JF5; 25-DEC-02.
PDB; 1JF6; 25-DEC-02.
PDB; 1J12; 18-DEC-02.
PDB; 1J1B; 12-AUG-03.
PDB; 1JL8; 18-DEC-02.
R InterPro; IPR006589; Alp_amy1_cat_sub.
R InterPro; IPR006047; Alpha_amy1_cat.
R InterPro; IPR004185; Glyco_hydro_13ig.
R InterPro; IPR007110; 19-like.
R Pfam; PF00128; alpha-amylase; 1.
R Pfam; PF02903; alpha-amylase_N; 1.
R SMART; SMO0642; Amy; 1.
W Carbohydrate metabolism; Hydrolase; Glycosidase; Calcium-binding;
3D-structure
T ACT_SITE 325 325
T ACT_SITE 354 354
T ACT_SITE 421 421
T METAL 143 143 CALCIUM.
T METAL 145 145 CALCIUM (VIA CARBONYL OXYGEN).
T METAL 148 148 CALCIUM.
T METAL 149 149 CALCIUM.
T METAL 169 169 CALCIUM (VIA CARBONYL OXYGEN).
T METAL 171 171 CALCIUM.
T METAL 171 171
T METAL 3 5
T STRAND 7 7
T STRAND 12 14
T STRAND 15 19
T STRAND 20 21
T STRAND 22 30
T TURN 31 32

FT STRAND 36 42
FT TURN 44 45
FT TURN 48 49
FT STRAND 53 62
FT STRAND 66 74
FT TURN 76 77
FT STRAND 80 87
FT STRAND 93 97
FT TURN 98 99
FT STRAND 100 102
FT HELIX 105 108
FT STRAND 111 113
FT TURN 119 120
FT HELIX 126 130
FT STRAND 133 136
FT HELIX 138 141
FT TURN 146 147
FT TURN 151 152
FT STRAND 153 153
FT TURN 158 159
FT TURN 164 165
FT STRAND 167 167
FT HELIX 172 185
FT TURN 186 186
FT STRAND 189 192
FT STRAND 196 196
FT STRAND 212 212
FT TURN 214 216
FT HELIX 219 230
FT TURN 231 233
FT STRAND 236 240
FT TURN 242 243
FT STRAND 244 244
FT TURN 247 248
FT HELIX 250 258
FT HELIX 259 261
FT TURN 263 264
FT HELIX 265 267
FT STRAND 268 268
FT STRAND 270 270
FT STRAND 283 283
FT STRAND 285 285
FT TURN 291 292
FT STRAND 294 296
FT TURN 298 299
FT HELIX 301 317
FT TURN 318 318
FT STRAND 321 324
FT TURN 325 326
FT HELIX 327 329
FT HELIX 332 345
FT TURN 347 348
FT STRAND 350 353
FT TURN 360 366
FT STRAND 370 373
FT STRAND 374 384
FT HELIX 385 386
FT TURN 391 402
FT HELIX 403 404
FT TURN 407 410
FT HELIX 411 412
FT STRAND 414 416
FT TURN 420 421
FT HELIX 425 428
FT TURN 429 431
FT HELIX 433 443
FT TURN 444 445
FT STRAND 449 453
FT TURN 454 455
FT HELIX 456 458
FT TURN 459 459
FT TURN 465 468
FT HELIX 476 478

```

CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; U29610; AAA93068.1; "-"
DR InterPro; IPR004000; Actin_like.
DR Pfam; PF00022; actin; 1.
DR SMART; SMO0268; ACTIN; 1.
DR PROSITE; PS01132; ACTINS_ACT_LIKE; 1.
DR Structural Protein; Cytoskeleton.
KW SEQUENCE 427 AA; 48635 MW; A53A0965B1E7EAF6 CRC64;
SQ
Query Match 5.1%; Score 89.5; DB 1; Length 427;
Best Local Similarity 20.6%; Pred. No.3;
Matches 64; Conservative 49; Mismatches 96; Indels 101; Gaps 17;
QY 25 DPYCAQXTCPC--TGSPIPVME---GDDIE--VRLQAPVWEFKYGDILGKLTMD 76
   || :|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||
Db 44 DPRRARRRMWCMPMAAGNIADLDPEFGDEAVENSKYQITMPV--RHQV----- 91
QY 77 AIGFRSTLTGKNYT-MEWYELFQIGNCTPFLR-----PMDAPFWCNQGAAC 123
   :|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||
Db 92 -----ENWTHMQF-----WEHCIFXRLCEPDHHFLLTEPLNAPENREYAEI 137
QY 124 FPEGIDVHAKKENGTLVQVATISGNMNCQAKWVKQDNNGIYYETM----- 170
   || :|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||
Db 138 MFE-----TNPVDPXYIAVQAVLALASWTSKQVTEKTLTGTVDSGDVTHVIPVAGY 192
QY 171 ----NVAKSPKGAE-TWF-----DSYDCKFVRLTFENKLA-EFGAEFKN 209
   :|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||
Db 193 VIGSIIHILAGRDITNFQLLRNENKIPRAETIEVAKRIETSYVCPDLVKEFK 252
QY 210 IETNYRIELFSGEPHYLGNETSVFPGTKN--TLGLAIRFYYP---FKPHLPTKEFL 264
   || :|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||
Db 253 YDTEPDKMF-----KTYEGIES-----VGKKPYNVGVYERFLPELFEFNDEIRSPFLT 302
QY 265 SLAQIFDAVI 274
   :|: || :|: || :|: || :|: || :|: || :|: || :|: || :|: ||
Db 303 PLPKVDETI 312
RESULT 5
DPOA HUMAN STANDARD; PRT; 1462 AA.
AC P09884;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA polymerase alpha catalytic subunit (EC 2.7.7.7).
GN POLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CX NCBI_TaxID=9606;
LN [1]
SEQUENCE FROM N.A.
RP MEDLINE=88196090; PubMed=3359994;
RA Wong S.W., Wahl A.F., Yuan P.-M., Arai N., Pearson B.E., Arai K.,
RA Korn D., Hunkapiller M.W., Wang T.S.-F.;
RT "Human DNA polymerase alpha gene expression is cell proliferation
RT dependent and its primary structure is similar to both prokaryotic
RT and eukaryotic replicative DNA polymerases.";
RL EMBO J. 7:37-47(1988).
LN [2]
SEQUENCE OF 1-8 FROM N.A.
RP MEDLINE=91172197; PubMed=2005899;
RA Pearson B.E., Nashener H.P., Wang T.S.;
RT "Human DNA polymerase alpha gene: sequences controlling expression in
RT cycling and serum-stimulated cells.";
RL Mol. Cell. Biol. 11:2081-2095(1991).
CC -1- FUNCTION: Polymerase alpha in a complex with DNA primase is a
CC replicative polymerase.
CC -1- CATALYTIC ACTIVITY: N deoxymucleoside triphosphate = N diphosphate

```



```

+ (DNA) (N).
-1- SUBCELLULAR LOCATION: Nuclear.
-1- MISCELLANEOUS: In eukaryotes there are five DNA polymerases:
alpha, beta, gamma, delta, and epsilon which are responsible for
different reactions of DNA synthesis.
-1- SIMILARITY: Belongs to the DNA polymerase type-B family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).

EMBL: X06745; CAA29920.1; -
EMBL: M64461; AAA52318.1; -
PIR: S00257; DJHDMC.
Genew: HGNC:9173; POLA.
GK: P09884; -
MIM: 312040; -
GO: GO:0005634; C:nucleus; NAS.
GO: GO:0003889; P:alpha DNA polymerase activity; NAS.
GO: GO:0006260; P:DNA replication; NAS.
InterPro: IPR006172; DNA_pol_B.
InterPro: IPR006134; DNA_pol_B_dom.
InterPro: IPR006133; DNA_pol_B_exo.
InterPro: IPR004578; Pol2.
Pfam: PF00136; DNA_pol_B; 1.
Pfam: PF03104; DNA_pol_B_exo; 1.
PRINTS: PR00106; DNAPOLB.
SMART: SM00486; POLBc; 1.
TIGRFAMs: TIGR00592; pol2; 1.
PROSITE: PS00116; DNA_POLYMERASE_B; 1.
Transferrase; DNA-directed DNA polymerase; DNA replication;
DNA-binding; Nuclear protein.
DNA BIND 650 715 POTENTIAL.
DNA BIND 1245 1376 POTENTIAL.
SEQUENCE 1462 AA; 165860 MW; 25C270B0A0DB38BE CRC64;

Query Match 5.0%; Score 88.5; DB 1; Length 1462;
Best Local Similarity 20.3%; Pred. No. 17;
Matches 71; Conservative 56; Mismatches 115; Indels 107; Gaps 20;

16 KRFDPRPKDXYCQAKYFPCPTG-----SPPVWEGD---DPIIV-----FRL 55
183 KRRTGASNPSPSVHTATAVPSGKTASPVSRKEPLTPTLRARFAGDVGVSSTEEBQ 242
56 QAPVEFKYGDLLGHUKIMHDAIGFRSTLTGNVTWMEWELFQNGCTEPHILRPMDAPF 115
243 ESGAMEFEGDF-----DERMEVEV-----DIEP-MAKA 272
116 WCNQGAACFEGSIDVHWKEN--GTLVQVATISGMENQAKW-VKQDNETGIYEETWN 172
273 WKES-----EPAEVKKQADBSGKTV-----SYLGSFLPDVSCWDIDGSSFSVQEVQ 323
172 VKASP---EKAGET-----WFDSD---CSKPYLRITNKLAEGAE-----FNKI 210
324 VDSHSLPLVKGADDEQVHFHYWLDADVDYQNPQGVVFLFGKWMISAEHVSQCVWVKNI 383
211 ENNYTRIFLYSGEPYLYGNETSVFEGTGKNT-LGLAKRFYYPFKPHLPTEKFLSLIQI 269
384 ERT-----LY-----FLPREMKIDINTGKETGTPISMKQVVEEPDEKATKKIKK- 429
270 FDVAIVHKOYFLFYNEFWFLPMKRPFIITYE---EIPDIPRNKTLIS 314
430 FKSKEVEKN---YAFELPDVPEKSEYLEVKYSALMPQIPQDLKGETS 474

ESUT 6
RUD_ARCFU STANDARD; PRT; 411 AA.
D TRUD_ARCFU
C 028596;

```

```

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable tRNA pseudouridine synthase D (EC 4.2.1.70) (Pseudouridylylate
DE synthase) (Uracil hydrolase).
GN TRUD OR AFI677.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RC MEDLINE=98049343; PubMed=9389475;
RA Klek H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kertavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman U.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370 (1997).
CC -1- FUNCTION: Could be responsible for synthesis of pseudouridine from
CC uracil-13 in transfer RNAs (By similarity).
CC -1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -1- SIMILARITY: Belongs to the pseudouridine synthase trnd family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

CC EMBL: AE000987; AAB89568.1; -
CC PIR: D69459; D69459.
CC TIGR: AF1677; -.
CC HAMAP: MF_01082; -.
CC InterPro: IPR001656; UPP0024.
CC Pfam: PF01142; UPP0024; 1.
CC TIGRFAMs: TIGR00094; TIGR00094; 1.
CC PROSITE: PS01268; UPP0024; 1.
KW tRNA processing; Lyase; Complete proteome.
FT ACT SITE 81
FT SITE 81
SQ SEQUENCE 411 AA; 47715 MW; 109B5250D969922E CRC64;

Query Match 5.0%; Score 88; DB 1; Length 411;
Best Local Similarity 20.3%; Pred. No. 3.8;
Matches 55; Conservative 45; Mismatches 97; Indels 74; Gaps 13;

44 MEGDDIEVFRLQAPVMEFKYGDLLGHUKIMHDAIG-----FSTLTGNVTWMEWEL 96
38 LSDSDPFLIRVEKKNM-----DTLNFARVLNAGISQKISFAGTQDKALTVQYPSI 92
97 FQNGCTEPHILRPMDAPFVCNCGAACFEGIDVDVHME-----NGTLVQVATISGN 148
93 YGVKK-----EELERVNLDAKIEVIGYARRALQDGLIGN 128
149 MENOAKWVKQDNETGIYEETWNVVASPEKGAETWF--DSYDCKFVLRTNKLAEGAE 206
129 FF-RIRVVGCRDGE--IFQETRN-ELMEKGPNNFQIGRFGSIRFIITHVEGKLL----- 178
207 FKNITNTYTRIF-LYSGEPYLYGNE-----TSVGEPTGNKTLGLAKRFYYPFKPHLP 260
179 ---LQNNYEAFVYVAKPFEGENEVEVKRIEILMETDARKLGLRELPRKYLRERNLQK 235

```

```

Y 261 -----EFLLSLQIFDAVIVHK-QEYLF 282
| | | | | : | | | | |
b 236 IREKSESEALSLPKNLKMFEHAAQSYIF 266

RESULT 7
100 HSV7J STANDARD; PRT; 755 AA.
C P52519;
T 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Large structural phosphoprotein homolog (PPI00).
U1.
S Human herpesvirus (type 7 / strain J1) (HHV7).
C Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C Alphaherpesvirinae; Simplexvirus.
X NCBI_TaxID=57278;
N [1]
P SEQUENCE FROM N.A.
A Nicholas J.;
L Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
C -1- SIMILARITY: TO THE LARGE STRUCTURAL PROTEOPROTEINS OF HSV-6 AND
HCV UN32.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
C EMBL: U43400; AAC54672.1; -
C F1R; T41912; T41912.
W Matrix protein, Phosphorylation.
Q SEQUENCE 755 AA; 86580 MW; 4083744CCG3F91DA CRC64;

Query Match
Best Local Similarity 20.7%; Score 87.5; DB 1; Length 755;
Matches 58; Conservative 37; Mismatches 98; Indels 87; Gaps 15;

Y 114 PF-WCNOGAAC-----FEEGIDD--VHMEN-GTLVQVATISGN-----MENQAKXV 157
| | | | | : | | | | |
b 7 PRAWISDEKCKLRSFENISSLPVDIRNPWILSQCVKXGNSINNKILYNMLIM- 65

Y 158 KODNETGIYETWNVKASPEKAEIWEISYDSKFEV-----LPTNKLA 202
| | | | | : | | | | |
b 66 -----IYHQTLCKKKDPD-EEVWQELIKYOKILKDYLEQROMITDYSSILTSFNKV- 116

Y 203 FGAEPKNIETNTYTRI--FLYSGEPYLGK-----ETSVFG--PTGNKTLGLAKRF 249
| | | | | : | | | | |
b 117 FETEFENVAKDKLIGSFLRWGVTHADYVNLTEEREIEBENOKAKNNMISFTTYI 176

Y 250 YYPFKKH-----LPTKEFLSLIQ-----TFDAVIVHKQFY 280
| | | | | : | | | | |
b 177 VDPWNNNGYVYNINRNLVILGNLITLHGSMMMETALNTINEKNALIKAIENNNKV 236

Y 281 LFVNFEPYFLPKK---PPFIKITYEIDPLIRNKTLISGL 316
| | | | | : | | | | |
b 237 SIYSYOILSLPLTSRVTSPFKILTEDEVDVITKSLIHAL 276

RESULT 8
HSD_EMENT STANDARD; PRT; 1184 AA.
C P78611; Q00744;
T 15-JUL-1999 (Rel. 38, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Chitin synthase D (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl
transferase D) (Class-V chitin synthase D).

```

```

GN CHSD OR CHSE.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiaceae; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 89;
RX MEDLINE=96285568; PubMed=8709948;
RA Motoyama T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
RT "The Aspergillus nidulans genes chsA and chsD encode chitin synthases
RT which have redundant functions in conidia formation."
RL Mol. Genet. 251:442-450(1996).
RN [2]
RP ERRATUM.
RX MEDLINE=97188530; PubMed=9037115;
RA Motoyama T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
RL Mol. Genet. 253:520-528(1997).
RN [3]
RP REVISIONS.
RA Motoyama T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RX MEDLINE=96406387; PubMed=8810520;
RA Specht C.A., Liu Y., Robbins P.W., Bulawa C.E., Iartchouk N.,
RA Winer C.R., Riggle P.J., Rhodes J.C., Dodge C.L., Culp D.W.,
RA Borja P.T.;
RT "The chsD and chsE genes of Aspergillus nidulans and their roles in
RT chitin synthesis."
RL Funct. Genet. Biol. 20:153-167(1996).
CC -1- FUNCTION: Plays a major role in cell wall biogenesis.
CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + {(1,4)-(N-acetyl-
CC beta-D-glucosaminyl)}(N) = UDP + {(1,4)-(N-acetyl-
CC beta-D-glucosaminyl)}(N+1).
CC -1- SUBCELLULAR LOCATION: Plasma membrane-bound.
CC -1- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS
CC V.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D83246; BAA1866.2; -
CC EMBL: U52362; AAA97482.1; -
CC InterPro: IPR004835; Chitin synth_fng.
CC InterPro: IPR001199; Cyt_B5.
CC Pfam: PF03142; Chitin_synth_2; 1.
CC Pfam: PF00173; heme_1; 1.
CC Transferase; Glycosyltransferase; Transmembrane; Cell wall;
CC Multigene family.
KW TRANSMEM 221 241 POTENTIAL.
KW TRANSMEM 476 496 POTENTIAL.
KW TRANSMEM 1039 1059 POTENTIAL.
KW TRANSMEM 1073 1093 POTENTIAL.
KW TRANSMEM 1097 1117 POTENTIAL.
SQ SEQUENCE 1184 AA; 133504 MW; D6157184F154ED5 CRC64;

Query Match
Best Local Similarity 4.9%; Score 87; DB 1; Length 1184;
Matches 75; Conservative 44; Mismatches 132; Indels 114; Gaps 20;

QY 33 TFCPGSPFVWEGDDIEVF-----RLQAPWEPKYGDLGHLKIMDAIGRS---TLT 85
| | | | | : | | | | |
DB 189 TYC---SVITWAPDFVLCFGMPKQKQSAWREKIG-LISIIIMIAFVCGLTGFGTAT 244
| | | | | : | | | | |
QY 86 -----GKNTTMEYELFQCGNCTFFHLR--PEN-----DAPF-WCNOGAAC 123
| | | | | : | | | | |

```

```

245 VCGTPPTBLKINIGSGYMIPIHQAYDLTKSTHPAAAGI PDMTNVLXDLPHKYGGQDSF 304
124 FEEGID-----DYHMKNGTLVQYATISGMFNOMAKVRODNETGIYETW 17C
305 FFQEVNAGCKGLITRTNSNDIPTNSGDLAWY--FPCIAFNODGS--SEPNITVSYNGM 36C
171 --NVKASPEKGAE-----TWFDSDYDCKSPVLTATFNKLAEPGAE-----PKN 20S
361 ACHTSGSARKSPYSLKNSGDYFTMEDTKNRS-----KLAVYSGVLDLNLNWFDD 413
210 IETNTRIF-----LYSGEPTIYI--NETSVFGPTGNKTLG-LAIKR 24E
414 TVNVPTFKXDLRDNDIRGVDTLYTFQTEGDKQIGKCLSLIKVGSIDTDTVGCIAQV 473
249 FYYPKPKHPLPTKEFLLSLQIFDAIVYHKQFLFYNFYFWLPMKPFPIKITIYEIRPI 30E
474 VLY-----VSILFIISTIVYKFAFALFQ--WFLAPFAAKTSMGAVSKA 51E
309 RNKTL 313
519 RNQOI 523

```

RESULT 9

3A2_MOUSE STANDARD; PRT; 254 AA.

```

28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
15-MAR-2004 (Rel. 43, Last annotation update)
3-oxo-5-alpha-steroid 4-dehydrogenase 2 (EC 1.3.99.5) (Steroid
5-alpha-reductase 2) (SR type 2) (5 alpha-SR2).
SRD5A2 OR 5ARR2.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
TISU0B=Kidney;
MEDLINE=21882004; PubMed=11884637;
Takeyama K., Kato S.;
"Transcriptional regulation of the mouse steroid 5alpha-reductase
type II gene by progesterone in brain.";
Nucleic Acids Res. 30:1387-1393 (2002).
-1- FUNCTION: Converts testosterone into 5-alpha-dihydrotestosterone
and progesterone or corticosterone into their corresponding 5-
alpha-3-oxosteroids. It plays a central role in sexual
differentiation and androgen physiology (By similarity).
-1- CATALYTIC ACTIVITY: A 3-oxo-5-alpha-steroid + acceptor = a 3-oxo-
delta(4)-steroid + reduced acceptor.
-1- SUBCELLULAR LOCATION: Integral membrane protein. Microsomal
intracellular membrane (By similarity).
-1- SIMILARITY: Belongs to the steroid 5-alpha reductase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB049456; BAB40179.1; -.
MGD; MGI:2150380; Str5a2.
GO; GO:0003865; F:3-oxo-5-alpha-steroid 4-dehydrogenase activity; IMP.
GO; GO:0003539; P:male genital morphogenesis; IMP.
GO; GO:0006694; P:steroid biosynthesis; IMP.
InterPro; IPR001104; Strd5A_dhc.
Pfam; PF02544; Steroid_dh; 1.
PROSITE; PS50244; S5A_RNDUCTASE; 1.
Oxidoreductase; Microsome; Transmembrane; Sexual differentiation;
Multigene family.

```

```

FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 146 166 POTENTIAL.
FT TRANSMEM 206 226 POTENTIAL.
SO SEQUENCE 254 AA; 28619 MW; 8D25019E8DC4DF47 CRC64;

```

Query Match 4.8%; Score 65; DB 1; Length 254;
 Best Local Similarity 27.3%; Pred. No. 3.8;
 Matches 35; Conservative 16; Mismatches 39; Indels 38; Gaps 7;

```

QY 197 FNKLAEFGAEFKNIETNY---TPIFLYSGE-PTYL-----GNETSVFGPTGNKTLGL 244
DB 27 FGKDAVYGHSESSVSSGVPLPARIAWPLQELPFSVSVGMIAWQPSRLFGPGNVLLGL 86
QY 245 AIKRFYYPKPKHPLPTKEFLLSL---QIFDAIVYHK-----QFIYFNFEY-- 287
DB 87 -----FSAHYHFKRTIYSLTRGRPLSAVIFLKATAPCIINGLLQAYVLYVCAEYPE 138
QY 288 -WFLPMKF 294
DB 139 EMTDMRF 146

```

RESULT 10

BPH1_YEAST STANDARD; PRT; 2167 AA.

```

P25356; Q02396; Q07348;
01-MAY-1992 (Rel. 22, Created)
10-OCT-2003 (Rel. 42, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Belge protein homolog 1.
BPH1 OR YCR032W OR YCR32W OR YCR591 OR YCR601.
Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=92116648; PubMed=1837415;
Rodriguez F., Martegani E., Mauri I., Alberghina L.;
"The sequence of 8.8 kb of yeast chromosome III cloned in lambda
PM2270 contains an unusual long ORF (YCR601).";
Yeast 7:631-641 (1991).
[2]
SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=92244356; PubMed=1574125;
Oliver S.G., van der Aart O.U.M., Agostoni-Carbone M.L., Aigle M.,
Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.F.G.,
Bent P., Berben G., Bergantino E., Bileau N., Boile P.-A.,
Botolin-Fukuhara M., Brown A.U.P., Brown R., Buhler J.-M.,
Cattigiani G., Chanet R., Contreras R., Crouzet M., Daignan-Bornier B.,
De Haan M., Defoor E., Delgado M.D., Demolder J., Dotra C., Dubois E.,
Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Fabre F.,
Fairhead C.A., Faye G., Feldmann H., Fiers W.,
Francinques-Gallard M.-C., Franco L., Frontali L., Fukuhara H.,
Fuller L.B., Gent M.E., Gigot D., Gilliquet V., Glansdorff N.,
Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haesemann M.,
Hateb D., Hegemann J.H., Heibert C.J., Hilger F., Hohmann S.,
Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
Jacc C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
Kleinhans U., Kreisel P., Lafranchi G., Lewis C., van der Linden C.G.,
Luchini G., Luczenkirchen K., Maat C., Mannheim P., Manzano M.E.,
Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
Messenguy F., Mewes H.-W., Moilemans F., Montagne M.A., Navas L.,
Newton C.S., Olson M.V., Paillet C., Panzeri L., Pearson B.M.,
Pera J., Philippsen P., Pierard A., Plantan R.J., Plevani P.,
Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
Sanz E., Schaffl-Gerstenschlaeger I., Scherens B., Schweitzer B.,
Shu Y., Skala T., Slonimski P.P., Sor F., Soustelle C.,
Spiegelberg R., Staveva L.I., Steensma H.Y., Steiner S., Thierry A.,

```

A Thiros G., Triano L.N., Urrestarazu L.A., Valle G., Vetter I.,
A van Vliet-Reedijk J.C., Volckaert G., Vreken P., Marmington J.R.,
A von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
A Zimmermann F.K., Sgouras J.G.;
T "The complete DNA sequence of yeast chromosome III";
L Nature 357:38-46(1992).
L [3]
P REVISIONS.
P Valles G., Volckaerts G.;
L Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
P [4]
P SEQUENCE OF 335-2167 FROM N.A.
P MEDLINE=91335897; PubMed=1870232;
X dia Y., Slonimski P.P., Herbert C.J.;
T "The complete sequence of the unit YCR59, situated between CRY1 and
T MAT, reveals two long open reading frames, which cover 91% of the
T 10.1 kb segment";
L Yeast 7:413-424(1991).
L [5]
C SEQUENCE OF 1760-2167 FROM N.A.
C MEDLINE=92133166; PubMed=1776366;
X Wicksteed B.L., Roberts A.B., Sagliocco F.A., Brown A.J.P.;
A "The complete sequence of a 7.5 kb region of chromosome III from
T Saccharomyces cerevisiae that lies between CRY1 and MAT";
T Yeast 7:761-772(1991).
C -1- FUNCTION: May be involved in acetic acid export.
C -1- SIMILARITY: Contains 1 BRACH domain.
C -1- SIMILARITY: Contains 4 WD repeats.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).
C -----
C EMBL; X62452; CAA44309.1; -
R EMBL; X58720; CAC42982.1; -
R EMBL; X59075; CAA41798.1; -
R EMBL; S78624; AAB21258.1; -
R PIR; S19444; S19444.
R Germline; 138938; -
R SGD; S0000628; BPH1.
R GO; GO:0009268; P:response to pH; IMP.
R InterPro; IPR000409; Beige_BRACH.
R InterPro; IPR001680; WD40.
R Pfam; PF02138; Beach; 1.
R ProDom; PD007848; Beige_BRACH; 1.
R PROSITE; PS50197; BEACH; 1.
R PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.
R PROSITE; PS50294; WD_REPEATS_REGION_1.
R PROSITE; PS50082; WD_REPEATS_2; FALSE_NEG.
R Repeat; WD repeat.
M REPEAT 1927 1965 WD 1.
M REPEAT 1976 2015 WD 2.
M REPEAT 2072 2111 WD 3.
M REPEAT 2129 2166 WD 4.
M DOMAIN 1545 1839 BEACH.
M CONFILICT 1893 1893 F->N (IN REF. 5).
M SEQUENCE 2167 AA; 250870 MW; ED7D607ADA982CD CRC64;
Query Match 4.7%; Score 83.5; DB 1; Length 2167;
Best Local Similarity 18.1%; Pred. No. 77;
Matches 48; Conservative 52; Mismatches 96; Indels 69; Gaps 9;
Jy 115 FMCQGAACPFEGDDVWKEKG-TLVQVATISGNMF---NQAKWYKODNE----- 162
Jb 1170 FNNYQVRFNITNTNIFKNNNSPISVRQIKNVYWKABEYVQNNKKCLILFRK 1229
Jy 163 --TGIYF-----TNVVASPKGAETWFDSDYCKEYLR-----TFNKLAERGA 205
Jb 1230 DNTSLDFKIKKISRYTYNLKTDRENNVAFYRNINILFIHKLTLEIQSNPNSSCKWSS 1289

Qy 206 EFKNIETWYTRIFLYSGEPTYL-----GNETSVPGTGNKTLGLAIKRFYPPKPLP 258
Db 1290 DAEDFDGKRRLPLAMEPKXEPLINBEDANDQITG--GNQRRSSGILSYEFLHME 1347
Qy 259 TKE-----FLSLQIPAVVHKQFYLPNFEYF 289
Db 1348 TLESSEPGDINENRKLRLKXNDISIAITWCSLIGLEIKGILIHGSNLYAFVSDYF 1407
Qy 290 LPMKPFKIITYEELPLPIRNKTL 314
Db 1408 SLEDKKILKLS--EVSQESRDWTVS 1430
RESULT 11
YCOL_ARCFU STANDARD; PRT; 228 AA.
ID YCOL_ARCFU
AC 029067;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF1201.
GN AF1201.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxId=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischman R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Arlrich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).
C -----
C EMBL; AB001021; AAB90048.1; -
DR PIR; H69399; H69399.
DR TIGR; AF1201; -
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 228 AA; 27209 MW; 3B29C4F593A26B68 CRC64;
Query Match 4.7%; Score 83; DB 1; Length 228;
Best Local Similarity 24.6%; Pred. No. 4.9;
Matches 29; Conservative 24; Mismatches 55; Indels 10; Gaps 5;
Db 169 TNVWVASPKGAETWFDSDYCKEYLR-TFNKLAERGAETWYTRIFLYSGEPTYL 226
Jy 5 SWNRSGEQR---FHTTSCVDYILRIEKKKFLCCKIKMKKTKVNRGTYVQDAYK 60
Qy 227 -LGNETSVFPGTGNKTLGLAIKRFYPPKPLP-----SLQIPDAVIVHKQFYL 281
Db 61 LIRLITELVYGNQSN-VLDQAIRLILYDVSKYKSKADRIILREKLINPENILISKQEM 117

```

ESULT 12
122 HUMAN STANDARD; PRT; 889 AA.
09E206; 09HCW1;
28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Purified alpha-mannosidase Clorf22 (EC 3.2.1.-).
C1ORF22.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=21218927; PubMed=11318611;
Sood R., Bonner T.I., Makalowska I., Stephan D.A., Robbins C.M.,
Comors T.D., Morgensesser S.D., Su K., Faruque M.U., Pinkett H.,
Graham C., Baxevanis A.D., Klinger K.W., Landes G.M., Trent J.M.,
Carpen J.D.,
"Cloning and characterization of 13 novel transcripts and the human
RG88 gene from the 1q25 region encompassing the hereditary prostate
cancer (HPC1) locus."
Genomics 73:211-222(2001).
[2]
SEQUENCE OF 16-469 FROM N.A.
Cobley V.,
Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
-1 SIMILARITY: Belongs to family 47 of glycosyl hydrolases.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@sib-sib.ch).
C
R EMBL AF288393; MAG60613.1; -.
R EMBL AL096819; CAC03447.1; -.
R HSSP; P32906; 1D12.
R Genew; HGNC:16787; Clorf22.
R InterPro: IPR000886; ER_target_S.
R InterPro: IPR001382; Glyco_hydro_47.
R InterPro: IPR003137; PA.
R Pfam; PF01532; Glyco_hydro_47; 1.
R Pfam; PF02225; PA; 1.
R PRINTS; PR00747; GLYHDRLASE47.
R ProDom; PD003239; Glyco_hydro_47; 1.
R PROSITE; PS00014; ER_TARGET_1.
R Hypothetical protein; Hydrolase; Glycosidase; Glycoprotein;
R Endoplasmic reticulum.
W SITE 886 889 PREVENT SECRETION FROM ER (POTENTIAL).
T CARBOHYD 75 75 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 468 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 767 767 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 771 771 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL).
Q SEQUENCE 889 AA; 100304 MW; 21354A62C5901666 CRC64;

Query Match 4.7%; Score 83; DB 1; Length 889;
Best Local Similarity 20.6%; Pred. No. 27;
Matches 65; Conservative 41; Mismatches 108; Indels 102; Gaps 17;

44 MEGDDIEVRPLQAPWEKYGDLLG-HDKIMHDAIGFSTLTGKNYTWENT-ELFOL 99
b 93 LNDVNVVSPETIRY---LGLLGHSLAI-----LKEGYMOWNDLDM 139
y 100 G-----NCTFPHLRPMMDAPFMCAGACFEGCIDVHWKENG 137
b 140 AKQLGYKLLPAFNTTGLPYPRINLKFKIRKPEARAGTETDTCTAC-----AG 187

```

```

QY 138 TLV-QVATIS-----GNMKNQAK-----WVKQDNETGIYETWNVKAS-----PEKGAET 182
Db 188 TLIEFALSLFTGATLIEEYARAKALDPLMEKQRSSNLVGTITINIRHGDWVRKSGVGA 247
QY 183 WEDSYDCKFVLRPNKLAERGAERKNIETNYTRFLYSGETYIGNTSVGFPGNKL 242
Db 248 GIDSY-YEYLLKAYVLGD-DSFLERENTHYDAIMRYISQPPPL-----L 290
QY 243 GLATKRFYPPKPHLPTEKEFLSLQLF-----DAVIHKQRYLPFYEPYFLP 291
Db 291 DVHI-----HKPMINATWMDALAFPPGLQVLKGDIRPALETHEMLYQVIK-KHNFILP 343
QY 292 MKPFP-IKITYEELPL 306
Db 344 EAFITDFRVHNAQHPL 359

RESULT 13
HGPB_HAEBIN STANDARD; PRT; 999 AA.
AC 087236;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin and hemoglobin-hapoglobin binding protein B precursor.
GN HGPB.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OC NCBI_TaxID=727;
[1]
SEQUENCE FROM N.A.
RC STRAIN=HI689 / Serotype B;
RA MEDLINE=98427137; PubMed=9746572;
RX Ren Z., Jin H., Morton D.J., Stull T.L.;
RT "hgpB, a gene encoding a second Haemophilus influenzae hemoglobin- and
RT hemoglobin-hapoglobin-binding protein."
RL Infect. Immun. 66:4733-4741(1998).
[2]
SEQUENCE FROM N.A.
RP STRAIN=ElA / serotype B;
RC Morton D.J., Stull T.L.;
RA "Conservation of hemoglobin/hemoglobin-hapoglobin binding proteins in
RT Haemophilus influenzae."
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
CC ADDITION OR LOSS OF CCA REPEAT UNITS WOULD CHANGE THE READING OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
R EMBL AF022910; AAC60790.1; -.
R EMBL AF259266; AAK51630.1; -.
R InterPro: IPR006970; PT.
R InterPro: IPR00531; Tonb_boxC.
R Pfam; PF04886; PT; 1.

```


BEST LOCAL SIMILARITY

Best Local Similarity 21.1%; Prod No. 33;
Matches 56; Conservative 33; Mismatches 87; Indels 90; Gaps 12.

Y 18 FDERPRDPYCQAKYTFCTGSPFPVMEGGDDIEVERLQAPVWEFKYDGLCHLKINMDA 77
||| |||
||| ||| : | : | :
C 53Z FDPLPKAD-----RLVDPI-----THVKLNNEG 554
||| |||
Y Y 78 -----IGRR--STLTGRKTYMENEYLPGNGCTEFLHLPRENDAPFWCNQG----- 120
||| |||
B 55S KLAGMLIQGFENVLSLNNKKRTLSGMSKLKYLVNMP-LQT-SSSEFENFESENNVNA 612
||| |||
Y 121 -----AACFGEGIDIV-----HWK---ENGTLVOATISGNFMNQAKWK 158
||| |||
B 613 EICTEVEVRLLPTCPAEDEBSIVNSGRKTOWHMHGCGQPGEALPDVIDLSMAEEKHLYK 672
||| |||
Y 159 QDNETGI-YYE-TWN-----VKASPEKAETWPSDYDCSKFYJRTFNKLAFGA 206
||| ||| : | : | :
T T 673 KEKGQGISPEAMTWVAQHPSPAVELAKEINGVALLEDYPDENGWLMYKKQLLINGFAH 732
||| |||
Y b 207 FKNIENITRIELYSGEPTYLGNETS 232
||| ||| : | : | :
P B-HUMAN LRDGTTSGMVLVVGVQMTKEKNOTA 758
||| |||
RESULT 15
PB_HUMAN STANDARD; PRT; 4563 AA.
D_ABP_HUMAN PO4I14; O00502; Q13787;
C C 01-NOV-1986 (Rel. 03, Created)
T T 01-NOV-1986 (Rel. 03, Last sequence update)
T T 15-MAR-2004 (Rel. 43, Last annotation update)
E Apolipoprotein B-100 precursor (Apo B-100) [contains: Apolipoprotein
E B-48 (Apo B-48)].
E N APOB.
S Homo sapiens (Human).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X NCBI_TaxId=9606;
X [1]
P SEQUENCE FROM N.A.
MEDLINE=87016385; PubMed=3763409;
X Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J.,
A Blackhart B., McCarthy B.U., Mahley R.W., Levy-Wilson B., Scott J.;
T "Complete cDNA and derived protein sequence of human apolipoprotein
B-100.";
L Nucleic Acids Res. 14:7501-7503(1986).
[2]
P SEQUENCE FROM N.A., AND VARIANT GIU-4181.
MEDLINE=88003974; PubMed=3652907;
X Ludwig E.H., Blackhart B.D., Pierotti V.R., Caiati L., Fortier C.,
A Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.U.;
T "RNA sequence of the human apolipoprotein B gene.";
L DNA 6:363-372(1987).
[3]
P SEQUENCE FROM N.A., AND VARIANTS ILE-98 AND GIU-4181.
MEDLINE=87008488; PubMed=3759943;
X Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H.,
A Gotto A.M., Jr., Chan L.;
T "The complete cDNA and amino acid sequence of human apolipoprotein
B-100.";
L J. Biol. Chem. 261:12918-12921(1986).
[4]
P SEQUENCE FROM N.A.
MEDLINE=87041416; PubMed=3464946;
X Law S.W., Grant S.B., Higuchi K., Hospattankar A.V., Lackner K.J.,
A Lee N.W., Brewer H.B., Jr.;
T "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and
derived amino acid sequence.";
L Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
[5]
P SEQUENCE FROM N.A.
MEDLINE=87161758; PubMed=3030729;
X Cladarias C., Hadzopoulou-Cladarias M., Nolte R.T., Atkinson D.,

RA Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,
 RA Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
 RA Lee F.-S., Gu Z.-W., Goto A.M., Jr., Chan L.,
 RT "Sequence, structure, receptor-binding domains and internal repeats
 RT of human apolipoprotein B-100.";
 RL Nature 323:738-742(1986).
 [15]
 RN CALCIUM-BINDING DATA.
 RP MEDLINE=86242245; PubMed=3087360;
 RA Dachtel N., Lee D.M., Mok T.,
 RT "Apolipoprotein B is a calcium binding protein.";
 RL Biochem. Biophys. Res. Commun. 137:493-499(1986).
 [16]
 RP PALMITOYLATION OF CYS-1112.
 RX MEDLINE=20143590; PubMed=10679026;
 RA Zhao Y., McCabe J.B., Vance J., Berthiaume L.G.,
 RT "Palmitoylation of apolipoprotein B is required for proper
 RT intracellular sorting and transport of cholesterol esters and
 RT triglycerides.";
 RL Mol. Biol. Cell 11:721-734(2000).
 [17]
 RP VARIANT SER-4338.
 RX MEDLINE=91071750; PubMed=1979313;
 RA Navajas M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,
 RA Cuny G., Cambien F., Roizes G.,
 RT "Detection by denaturing gradient gel electrophoresis of a new
 RT polymorphism in the apolipoprotein B gene.";
 RL Hum. Genet. 86:91-93(1990).
 [18]
 RP VARIANT FDB GLN-3527.
 RX MEDLINE=89098975; PubMed=2563166;
 RA Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
 RA McCarthy B.J.,
 RT "Association between a specific apolipoprotein B mutation and
 RT familial defective apolipoprotein B-100.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
 [19]
 RP VARIANT LEU-2739.
 RX MEDLINE=91016974; PubMed=2216805;
 RA Huang L.-S., Gavish D., Breslow J.L.,
 RT "Sequence polymorphism in the human apob gene at position 8344.";
 RL Nucleic Acids Res. 18:5922-5922(1990).
 [20]
 RP VARIANT FDB CYS-3558.
 RX MEDLINE=95190020; PubMed=7883971;
 RA Pullinger C.R., Hennessy L.K., Chatterton J.E., Liu W., Love J.A.,
 RA Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.,
 RT "Familial ligand-defective apolipoprotein B. Identification of a new
 RT mutation that decreases LDL receptor binding affinity.";
 RL J. Clin. Invest. 95:1225-1234(1995).
 [21]
 RN VARIANTS LEU-1437; SER-1914; LYS-2566; THR-3121; ALA-3945; MET-4128
 RP AND THR-4481.
 RX MEDLINE=97044521; PubMed=8889592;
 RA Poltier O., Ricard S., Behague I., Souriau C., Evans A.E.,
 RA Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.,
 RT "Detection of new variants in the apolipoprotein B (apo B) gene by
 RT PCR-SSCP.";
 RL Hum. Mutat. 8:282-285(1996).
 [22]
 RP VARIANTS FDB GLN-3527 AND CYS-3558.
 RX MEDLINE=97403938; PubMed=9259199;
 RA Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,
 RA Krempf M., Giraudet P., Junien C., Boileau C.,
 RT "Familial ligand-defective apolipoprotein B-100: simultaneous
 RT detection of the ARG3500->GLN and ARG3531->CYS mutations in a
 RT French population.";
 RL Hum. Mutat. 10:160-163(1997).
 [23]
 RN VARIANTS SER-1914; ARG-1923; LEU-2739; ASP-3319; THR-3427; GLN-3432
 RP AND ILE-3921.
 RX MEDLINE=98141125; PubMed=9490296;
 RA Ieren T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.,

RT "Screening for mutations of the apolipoprotein B gene causing
 RT hypocholesterolemia.";
 RL Hum. Genet. 102:44-49(1998).
 CC -1- FUNCTION: Apolipoprotein B is a major protein constituent of
 CC chylomicrons, VLDL and LDL. It functions as a recognition signal
 CC for the cellular binding and internalization of LDL particles by
 CC the apoB/E receptor.
 Query Match 4.7%; Score 83; DB 1; Length 4563;
 Best Local Similarity 19.7%; Pred. No. 2.2e+02;
 Matches 56; Conservative 39; Mismatches 123; Indels 66; Gaps 11;
 QY 44 MEGDDIEVRLQAPWVEFYKGDLLG--HLKIMHDAIGFSTLTGKNTMWEYELPOLG 100
 Db MDDEDDDS-----KMFYYSPOSSPDKLTIFKTELARRESDDEQIVNMEEAASG 4065
 QY 101 NCTFPHLRPEMDAPFWCNGACAPFEGIDDVHKENG-TLVQVATTSGNNPNOMAKVYQ 159
 Db 4066 ILT--SLKDNVP-----KATGVLYDVNKKYHWEHGLTLREVSSKLRLQNNAEWVYQ 4117
 QY 160 DNENGIYETWNYKASPEKGAETWPDSDSKVLRFTFNKLAEPGAFKNIETN-YTRIF 218
 Db 4118 ----GAIROIDDIDVRFQKASGTTGYQ-----EMKOKAQNLYGELL 4156
 QY 219 LYSGEPTYLGNETSVEFGPTGNKTLGLAIK-----RYYPERKPHLPYKEFL 263
 Db 4157 TQSGQASFOGLKONVPEGLVRYVQKHEMKVKHLIDSLIDPLNPRFQFGKPGIYTRTEL 4216
 QY 264 LSLIQIDDAVIYHKQFY-----LFYNEFWELPMKPEFLK 238
 Db 4217 CTMF-IREVGTVLSQVYSKVHNGSEILFSYFQDLVITLPEELRK 4259

Search completed: June 8, 2004, 14:04:39
 Job time : 14.1667 secs


```

b 207 YDCSNFVLRITKKAIEFGTEFKKLEWNTKFLVSGSEIYIGNETSLFGPGKNTLALAI 266
b 247 KRFFYPKPEHLPTKEFLSLQIFDAVIVHKQFLYFVFWFVPMFMPFKITYEELPL 306
b 267 KKFYGPFPFVYSTDFLNMFLKIDTVIINHOFILFNFWEFLPMKPPFKITYEETPL 326
b 307 PIRNKTLISGL 316
b 327 PTRHTTFDTL 336

RESULT 2
b 08C054 PRELIMINARY; PRT; 322 AA.
b 08C054;
b 01-MAR-2003 (Tremblrel. 23, Created)
b 01-MAR-2003 (Tremblrel. 23, Last sequence update)
b 01-JUN-2003 (Tremblrel. 24, Last annotation update)
b Similar to CER01D-LIPOUSCINOSIS neuronal protein 5 (Fragment).
b CINS.
b Mus musculus (Mouse).
b Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
b Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
b NCBI_TaxID=10090;
b [1]
b SEQUENCE FROM N.A.
b STRAIN=C57BL/6J; TISSUE=Olfactory brain;
b MEDLINE=22354683; PubMed=12466851;
b THE FANTOM Consortium,
b the RIKEN Genome Exploration Research Group Phase I & II Team;
b "Analysis of the mouse transcriptome based on functional annotation of
b 60,770 full-length cDNAs."
b Nature 420:563-573 (2002).
b EMBL: AK032293; BAC27797.1; -.
b MGI: 2442253; CINS.
b MGI: 2442253; CINS.
b NCBI_TaxID=10090;
b SEQUENCE 322 AA; 37338 MW; 04FE01B05BDEDB8A CRC64;

Query Match 80.7%; Score 1418; DB 11; Length 322;
Best Local Similarity 79.0%; Pred. No. 8e-116;
Matches 245; Conservative 30; Mismatches 35; Indels 0; Gaps 0;

```

```

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szatranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum."
RL Nature 418:79-85 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC116305; AA052280.1; -.
KW Hypothetical protein.
SQ SEQUENCE 180 AA; 20631 MW; 2DEB686F4FEC11AC CRC64;

Query Match 13.4%; Score 235; DB 5; Length 180;
Best Local Similarity 35.8%; Pred. No. 1.2e-12;
Matches 54; Conservative 25; Mismatches 60; Indels 12; Gaps 6;

QY 22 PRKDP-YCOAKY--TCPTGSPFP--VMESDDIEVFLQAPVWEKYGILLHLKIMED 76
Db 23 PNDNPFLLCOQRIQREDPCQ-TVPVWGTFTNDNDIEVYVYQAPVFEVFGNPFKLGYS 81
QY 77 AIGFRLTLGKXNYTWMEYELFQVGNCTFPHL--RPMNDAPFWGNOGACPFEGIDVHWK 134
Db 82 AIGFYLTITGLANTAYADAFEEVGNGLTLPNITVWGKDELMLWAGILCTVPYINETYWD 141
QY 135 EN----GLIVQVATISGNMFMQAKVWKODN 161
Db 142 KNITSTASKYWGTINGTQLONYIEMWQYN 172

RESULT 4
Q86JG6 PRELIMINARY; PRT; 378 AA.
ID Q86JG6;
AC Q86JG6;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szatranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum."
RL Nature 418:79-85 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC116982; AA051609.1; -.
KW Hypothetical protein.
SQ SEQUENCE 378 AA; 42306 MW; 2CF33D60851FD036 CRC64;

Query Match 11.5%; Score 202; DB 5; Length 378;
Best Local Similarity 23.4%; Pred. No. 2.4e-09;
Matches 65; Conservative 53; Mismatches 130; Indels 30; Gaps 8;

```

```

QY 48 DDIEVEFLQAPWEFFKGGDILGHLKIMHAIIFERSTLGNKNTMEWELPOLGCTEPHL 107
Db 25 DEIVLYLLEAPLRFSEKFGILLANVANAFHSGLGFYSLSNSETTYQVDFAVPGVLESILPN 83
QY 108 RPEMDAPFCWGNQACAFCEGIDIV-HKENGTLVQVATISGMENQAKWAKQDNETGIY 166
Db 84 KIENGELINMTSKVOFSDSIDNTGYSWE--KEIMTINNEIFETFLCMAQNNEISLY 140
QY 167 YETWNVKASPEKGAEMTSDYCSKRVLTENTKLELFGAEF-----KNIENYTRIF 218
Db 141 YQLEFDV--LNKRNTELLKVSVCNDFPMAGFALNLOGKATINOSILASRIYLLYLR-- 196
QY 219 LYSGEPTLYGNETSVEFGTKTLGLAKRFYFPPKHPPTKEFLLSLQIFDAVVIYHKQ 278
Db 197 -----BSIEYVODGSTNIEY-ISFYEAMUNISIPKQNGINGIELDE-LPKGN 243
QY 279 FYLFNPEYFELPMKRPFIKITYEELPIPIRKUTLSGL 316
Db 244 FYLYNGDIYNNVTLDSPEFSAISESPLIGERDLSNT 281

```

RESULT 5	OSNITH1	PRELIMINARY;	PRT;	633 AA.
ID	OSNITH1			
AC	OSNITH1			
DT	01-OCT-2002 (TREMBlrel. 22, Created)			
DT	01-OCT-2002 (TREMBlrel. 22, Last sequence update)			
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)			
DS	Putative secreted metalloprotease 3.			
GN	MEP3.			
OS	Trichophyton rubrum.			
OC	Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;			
OC	Omyceliales; Arthrodermataceae; mitosporic Arthrodermataceae;			
OC	Trichophyton.			
OX	NCBI_TaxID=5551;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Capoccia S., Lechame B., Zaung C., Monod M.,			
RT	"Trichophyton rubrum encoding metalloprotease.",			
RL	Submitted (Aug-2001) to the EMBL/Genbank/DDb databases.			
DR	EMBL: AF407189; ANNO3640.1; -;			
DR	EMBL: AF407190; ANNO3641.1; -;			
DR	GO: GO:0005576; C:cytosol; IEA.			
DR	GO: GO:0004322; F:metalloendopeptidase activity; IEA.			
DR	GO: GO:0008270; F:zinc ion binding; IEA.			
DR	GO: GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	InterPro: IPR001842; Peptidase M36.			
DR	InterPro: IPR006025; Pept M Zn BS.			
DR	Pfam: PF02128; Peptidase M36; 1.			
DR	PRINTS: PR00099; FUNGALYSIN.			
DR	PROSITE: PS00142; ZINC_PROTEASE. 1.			
KW	Metalloprotease; Protease.			
SO	SEQUENCE 633 AA; 69778 MW; 44DEFD7D8A01B92B6 CRC64;			

Query Match	5.7%;	Score 101;	DB 3;	Length 633;
Best Local Similarity	22.8%;	Pred. No. 3.3;		
Matches	61;	Conservative	34;	Mismatches 78;
				Indels 94;
				Gaps 17

```

QY 56 QAPVMEFKGDLGLKLTMHAI-----GFSSTLGGKRYMWEVLEFQLGNCFFPH 107
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 141 RAEV--EKRRDFSGMRAPFHGACKALNLPINADKTI---QTMNEHVEVFGT----- 187
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 108 REMDAPEWCGNAGACFEFGIDVHWKXNGTLV--QVATISGMF----- 150
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 188 SGAMSDP--QGKCYMA-----KEDGTLATWAEVETMDMDNMLLSYDAKETDKVH 236
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 151 -----NMAKAVQDNBTG---IYETWVVKASPEKGAETMPEDSDSKFVLATPE 197
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 237 NVVDVYSHATYQVYFWPIPDPTESGRREIVENPMILKTSF--FTWISDGNTVYTTTRGN 292
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 198 NKLAEFGAEFKNIETNRYRIFLYSGEPTLYGNETSVEGPTGNLTGLAIKRFYEPKPH 257
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

Db 293 NALAQN-----FDGGEDYLAN---YRP-NSKUL-----KEEIPYAPNM 327

QY 258 -PTKEFL-LSLLQIF-DAVIYHKQFYL 281
| | : | : | | | : | :
328 SPEKSYIDASYQLFYSANIYHDLTYM 354

Db

RESULT 6			
Q8A3V2			
ID	Q8A3V2	PRELIMINARY;	PT; 514 AA.
AC	Q8A3V2;		
DT	01-JUN-2003 (TEMBirel. 24, Created)		
DT	01-JUN-2003 (TEMBirel. 24, last sequence update)		
DT	01-OCT-2003 (TEMBirel. 25, last annotation update)		
DE	Putative beta-xylosidase.		
GN	B72852.		
OS	Bacteroides thetaiotaomicron.		
OC	Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;		
OC	Bacteroidaceae; Bacteroides.		
OX	NCBI_TaxID=818;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VPI-5482 / ATCC 29148;		
RX	MEDLINE=22550858; PubMed=12663928;		
RA	Xu J., Buzselli M.K., Himrod J., Deng S., Carmichael L.K.,		
RA	Chiang H.C., Hooper L.V., Gordon J.I.;		
RT	"A genomic view of the human-Bacteroides thetaiotaomicron symbiosis";		
RL	Science 299;2004-2076(2003).		
RL	EMBL: AE016937; AAO7958.1;		
DR	GO; GO:0004553; F1;hydroalase activity, hydrolyzing O-glycosyl . . .; IEA		
DR	GO; GO:0005975; P;carbohydrate metabolism; IEA.		
DR	InterPro; IPR006710; Glyco_hydro_43.		
DR	Pfam; PF004616; Glyco_hydro_43; 1.		
KM	Complete proteome.		
SO	SEQUENCE 514 AA; 5815 MW; 6F6E86653CC984B8 CRC64;		

Query Match Similarity	5.7%;	Score 99.5;	DB 16;	Length 514;
Best Local Similarity	21.8%;	Pred. No.3.4;		
Matches 52; Conservative	30;	Mismatches 92;	Indels 65;	Gaps 9

QY 11WVVPKRFDFRPRPCQAYKTECPTSPSPVW--GGDDIEVFRVLAQAPWEKCYKD-- 66
|||
|||
Db 294 WVIIVGLVDIMNGIGEPV--KWTIKRNTGKAPVVSFPGGSDFSDELNLQ-WQPNHPSD 350
-----LGLHLKIMH-----DAIGFSTLTGKNYTWY-----LFOL 99
67
Db 351 ADMNLTERRGWLILKATKADLRASRNMLTQKCIIGYGLVTTEDMDSWTEGGRAHFCI 410
|||
|||
QY 100 GNCTPPLRLREMDAPFWNGGACAFPEEI-----DYHWKENGTLVAYATISGMEN 151
|||
|||
Db 411 GN-----LFGNIGILKENGKNVLYLENGSGVEKVKVSGSKIIY 448
|||
|||
QY 152 QAKAKWKODNETGYIYETFWNKASP-EKGAETWEDSDSCGFVLRPNKLAEGAEKRN 209
|||
|||
Db 449 PPAITNAATNQHQLYSTDNKNLFFPCGBAYSJLRFGDWKARVGLYSINTLRDGNAFEN 507

RESULT 7		
Q8NB36		
ID	Q8NB36	PRELIMINARY; PRT; 682 AA.
AC	Q8NB36;	
DT	01-OCT-2002 (TrEMBLrel. 22, Created)	
DT	01-OCT-2003 (TrEMBLrel. 22, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Hypothetical protein NT2RP202760.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,	
RA	Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,	

RA Nagahari K., Sugano S., Isogai T.;
 RT "HRI human cDNA sequencing project";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL; AK075480; BA011645.1; -
 DR GO; GO:0016021; C:Integral to membrane; IEA.
 DR GO; GO:0005215; F:Transporter activity; IEA.
 DR GO; GO:0006810; P:Transporter; IEA.
 DR InterPro; IPR007114; MFS.
 DR InterPro; IPR005828; Sub transporter.
 DR InterPro; IPR005828; Sug_transporter.
 DR InterPro; IPR005988; SV2.
 DR Pfam; PF00083; Sugar_tr; 1.
 DR TIGRFAMs; TIGR01299; synapt_SV2; 1.
 DR PROSITE; PS00850; MFS; 1.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
 DR Hypothetical protein, Transmembrane.
 KM SEQUENCE 682 AA; 7660 MW; DIA76C115424CCTF CRC64;
 SQ

Query Match 5.6%; Score 98; DB 4; Length 682;
 Best Local Similarity 22.1%; Pred. No. 6.6;
 Matches 62; Conservative 40; Mismatches 104; Indels 74; Gaps 16;

QY 86 GKNYTMWYELFQGNCTPEH-----LRPMDAP-FKCNQG---AACFEGIDVHWK 134
 DB 325 GSAYQFHSWYFVL-VCAFPVFAIGALTTPESPREFLENGKDEAMVILKQYHDNMR 383
 QY 135 ENGLTVQVATISG-NMFQMAKWKVQDNETGIYETWNVKASPEKGAETWDSYDCKEV 193
 DB 384 AKGPERFVSYTHITTHQEGHLEIQSDTGTWQRMVRL-SLGGVGNFLASC--- 438
 QY 194 LRTEFKLAEFGAEFNIEIN---YTRIFLYSGEPTL-----GNESVF-- 234
 DB 439 -----FGPEYRITLMMGVFTMSFSYGLTWPEPMIRHLQAVDASRTKXVPPG 489
 QY 235 ----GPTGNKTLGLAIKRFYFEPK-----HLPTKEPLISLQ--IFDAVIHKQFY--- 280
 DB 490 EEVHEVTEFTLENDIHRGQYFNDKFLGLRLKSAFEDSLFEECYFEVYSSNAFFPNC 549
 QY 281 ----LEYN---FEYWFLPMKFFIKITY---EEIPUPI 308
 DB 550 TEINTVFINTDLFEYKFVNSR--LINSTLHNKEDCPIDV 587

RESULT 8

Q81AUS PRELIMINARY; PRT; 804 AA.

ID Q81AUS
 AC Q81AUS;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 23, last annotation update)
 DE Cullin-like protein, putative.
 GN PF08_0094.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL844507; CAD51265.1; -
 DR GO; GO:0007049; P:cell cycle; IEA.
 DR InterPro; IPR001373; Cullin.
 DR Pfam; PF00888; Cullin; 1.
 DR SMART; SM00182; CULLIN; 1.
 DR PROSITE; PS50069; CULLIN_2; 1.
 SQ SEQUENCE 804 AA; 97162 MW; 6B6E916BA99EAD2C CRC64;

Query Match 5.5%; Score 97.5; DB 5; Length 804;
 Best Local Similarity 19.8%; Pred. No. 9;
 Matches 57; Conservative 51; Mismatches 115; Indels 65; Gaps 14;

QY 23 KDDPYQQAATCTCPFGSP-----PVMEGDDIEVFR-QAPWWEKY-----GD 66
 DB 57 KKNPCYSKEVYRKXGSELSMYTIDKIKPLKNSDELNKTILIDAM-FKYSFYTKMNKK 115
 QY 67 LIGHIKIMIDAGFSTLTKKNTWMEYELFQGNCTPEHLEPMDAPFCWQAACFPE 126
 DB 116 FRLYLDRIY--VEYSSLSLSAVTNIIFI-----TLFNELR-----E 151
 QY 127 GIDDVHWK-ENGLTVQVATISGMENQMAKWKVQ-DNEG-IYF-----TWNVKAS 176
 DB 152 DIKNIIVELNNRLGEEIDQKELFCNIVEILYKELDNESNKKYEHDIKKIVENNNFY 211
 QY 177 EKCAETWDSYDCKEVLTENKLAEGFAEFNKIEN-----YTRIFLYSGEPT 226
 DB 212 KKKAEEMINDYPPDDYIISENAIEKEYEKNKSLNNDTCEKVTNIIVKILLYEKINTL 271
 QY 227 LGNETVVEGPTGNKTLGLAIKRFYFEPKPHLPTKEPLISLQIFDAVI 274
 DB 272 IDNKNVIFHLKNNNLS-SLRRTIIFS-YFP--EALTGKILIGEVY 315

RESULT 9

O81DS4 PRELIMINARY; PRT; 508 AA.

ID O81DS4
 AC O81DS4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
 DE Hypothetical protein.
 GN PF13_0225.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B., Denard N., Clark L., Line A., Barron A., Corton C.,
 RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.,
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL844509; CAD52545.1; -
 KM Hypothetical protein.
 SQ SEQUENCE 508 AA; 58842 MW; F563E34C1A9F25D2 CRC64;

Query Match 5.5%; Score 96; DB 5; Length 508;
 Best Local Similarity 20.0%; Pred. No. 6.7;
 Matches 52; Conservative 33; Mismatches 95; Indels 80; Gaps 11;

QY 126 EGIDVHWKENGTLVQVATISGMENQMAKWKVQDNETGIYF-----TWNVKASPEKG 179
 DB 121 DNNNNNNPHNNNNNNYNNVYTLDNKMPMKSKSMNDPNSNITNMINDGIMNKKEYPSNI 180
 QY 180 AETWDSYDCKEFLRTFN-----KLAEGFAEFNKIETN-YTRIFLYS 221
 DB 181 EN--YDYVDLNPFPQVNNQVLSANNIYVPLGRNVIQIENDSINGVNTNEYIKGTHV 238
 QY 222 GE-----PTL-----GNESVFGPTGNKTL 242
 DB 239 GRKFPNVLAGCOLPSLCTPSYTSNNNNNNNNNNNNNNNNNNNNNNNNNNIIGPFPNKTL 298
 QY 243 GLAI-----KRYVFEKPHLPTKEPLISLQIFDAVIHKQFYLFNFEYFLPMKFFP 296
 DB 299 NMEYIINDQKKEYIYF-----LNKGMNELTNDYNNWSTVK--FISPNVEYLMNTNPSPY 351
 QY 297 IKITY--EETPLPIRNK-TLS 314
 DB 352 YMLNINKEKLDLFFKKTLLS 371

RESULT 10

O9ZV89 PRELIMINARY; PRT; 359 AA.

ID O9ZV89
 AC O9ZV89;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)

```
01-MAY-1999 (TRENBLREL. 10, last sequence update)
01-JUN-2003 (TRENBLREL. 24, last annotation update)
F9K20.25.
F9K20.25.
Arabidopsis thaliana (Mouse-ear cress).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
STRAIN=cv. Columbia;
Vystotskaya V.S., Schwartz J.R., Toriumi M., Yu G., Li J., Liu S.,
Kremensskaya I., Iurov J., Araujo R., Buehler E., Conway A.B.,
Dewar K., Feng J., Kim C., Li Y., Shinn P., Davis R.W., Hocker J.R.,
Fiederspiel N.A., Theologis A.,
"Arabidopsis thaliana chromosome 1 BAC F9K20 sequence."
Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
STRAIN=cv. Columbia;
Theologis A.;
Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
EMBL; AC005679; AAC83039.1; -.
PIR; A96816; A96816.
InterPro; IPR004253; DUF231.
Pfam; PF03005; DUF231; 1.
SEQUENCE 359 AA; 40568 MW; A9887DD29786278B CRC64;

Query Match 5.4%; Score 95.5; DB 10; Length 359;
Best Local Similarity 21.5%; Pred. No. 4.7;
Matches 53; Conservative 39; Mismatches 97; Indels 57; Gaps 12;

32 YTEGCTGPIPMEDDDIEVRLQAPVMEFKYGGDLHLKIMHDAIGFSTLTAKNTM 91
80 YRMQPTGDIIPENGRDPLTRPKGKILF--VGDSLSN-----N 116
92 EWEYELFQNGCTFPHLRPEMDAPFWCNOGACFF--EGIDVHKENGTLVQVAT-- 144
117 MWVSL-----SCMLHAAYVNAKTFQANKGLSTFTPEGT--SYNLKNGFLVDLSGTR 171
145 ---ISGNFNOAKAVKQDNETGITYETNNVAKSPKGAETWFDSDSKFVLRTFNKLA 201
172 GLIKLIDISIRGNOMLISDVAILFNTFWHW---SHTGRAKTW-DYFGCDKIVKEMNME 226
202 EFGAEF-----KNLETNTRIELFVSGEPYTL-GNITSVFGPGAKTLG--LAIKRFY 250
227 AKTALITWMSKIDHNIDPSKTRVYGVSVPHLNGS---WGKPKGKTCLGETVYVQGPS 283
251 YPFKPH 256
284 YPGRPN 289

SUPT 11
3785 PRELIMINARY; PRT; 836 AA.
013785:
01-NOV-1996 (TRENBLREL. 01, Created)
01-NOV-1996 (TRENBLREL. 01, last sequence update)
01-JUN-2003 (TRENBLREL. 24, last annotation update)
APOB protein (Fragment).
APOB.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=86042646; PubMed=2932736;
Wei C.F., Chen S.H., Yang C.Y., Marcel Y.L., Milne R.W., Li W.H.,
Sparrow D.T., Goto A.M., Jr., Chan L.;
"Molecular cloning and expression of partial cDNAs and deduced amino
```

```
RT acid sequence of a carboxyl-terminal fragment of human apolipoprotein
RT B-100."
BL Proc. Natl. Acad. Sci. U.S.A. 82:7265-7269(1985).
EMBL; M12413; AA51742.1; -.
DR PIR; A27850; LPHUB.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000531; T0NB_BoxC.
DR PROSITE; PS00430; T0NB_DEPENDENT_REC_1; 1.
FT NON TER 1
SQ SEQUENCE 836 AA; 96317 MW; 5298C27EE21AB140 CRC64;

Query Match 5.4%; Score 95.5; DB 4; Length 836;
Best Local Similarity 20.5%; Pred. No. 14;
Matches 61; Conservative 39; Mismatches 125; Indels 73; Gaps 12;

37 TGSPIPV-----MEGDDIEVRLQAPVMEFKYGGDLG--HLKIMHDAIGFSTLTG 86
272 TSAPSPAVGTGMVDDEDDDS-----KNMFTYSSPSDPDKLTITKTELRVRESDEE 324
87 KNYTWWEYELPOLGNGCTFPHLRPEMDAPFWCNOGACFFEGIDVHKENG-TLVQVATF 145
325 TQIKNMWEBAASGLT--SLKDNVP-----KATGVLYDVNKKYHMEHTGLTAEVSK 376
146 SGNNFNOAKAVKQDNETGITYETNNVAKSPKGAETWFDSDSKFVLRTFNKLAIFGA 205
377 LRRNLQDHAEWYQ---GAIRETIDIDERFQKASGTGTGYQ----- 415
206 EKNLETN-YRIFLYSGEPYTLGNETSVPFGTNGKTLGLAIK-----RF 249
416 EFKDKAQLYELILTOBQASFGILKQNVFDGLVAVTQFFMKKHLIDSLIDFNFRF 475
250 YPFKPHLPTKEFLLSLQIFDAVIVHKQFY-----LFYNEEYVLPKPPFIK 298
476 QPFGKPGIYTRBELCTWR-IREVGVLSQVYSKVHNGSELLFSYQDVLITLPFLRK 532

RESULT 12
09VPL9 PRELIMINARY; PRT; 5322 AA.
ID 09VPL9; Q9N164;
AC 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-OCT-2002 (TRENBLREL. 22, last sequence update)
DT 01-OCT-2003 (TRENBLREL. 25, last annotation update)
DE CG3696 protein (KISMET-L long isoform).
GN KIS OR CG3660 OR CG3696 OR CG18326.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baas A., Baxendale U., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan P.R., Bouck J., Brocstein P., Broctlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Douc P.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiza S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
```

Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman J.T., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy D., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svitzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
Williams S.M., Moodale T., Morley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"the genome sequence of *Drosophila melanogaster*";
Science 287:2185-2195(2000).
[2]
SEQUENCE FROM N.A.
Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,
Carlson J.W., Center A., Chapple M., Davenport L.B., Dietz S.M.,
Dodson K., Dorey V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
Fertler A., Frise E., Galle R.F., Gang N.S., George R.A.,
Gonzalez M., Haegele J., Hoskins R.A., Hostin D., Howland T.J.,
Ileguam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,
Pacleb J., Pargass V., Park S., Patel S., Pfeiffer B.,
Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
Stapleton M., Strong R., Svitzkas R., Tector C., Tyler D.,
Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
"sequencing of *Drosophila melanogaster* genome";
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
Mitra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
Tupy J.L., Bergman C., Bertman B., Carlson J.W., Celinker S.E.,
Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,
Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
Aeburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
"Annotation of *Drosophila melanogaster* genome";
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
SEQUENCE FROM N.A.
Flybase;
Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
[6]
SEQUENCE FROM N.A.
Therrien M., Morrison D.K., Wong A.M., Rubin G.M.;
"A Genetic Screen for Modifiers of a KSR-Dependent Rough Eye Phenotype
in *Drosophila*";
Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
EMBL; A5003590; AAF51527.3; -
EMBL; AF215703; AAF43004.1; -
HSSP; P23197; IAP0.
FlyBase; Fgn0001309; Kis.
GO; GO:0000785; C:chromatin; IEA.
GO; GO:0005634; C:nucleus; IEA.
GO; GO:0005524; C:ATP binding; IEA.
GO; GO:0008026; F:ATP dependent helicase activity; IEA.
GO; GO:0003682; F:chromatin binding; IEA.

DR GO:0016787; F:hydrolase activity; IEA.
DR GO:0006333; P:chromatin assembly/disassembly; IEA.
DR InterPro: IPR000953; Chromo.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00385; chromo; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00176; SNF2_N; 1.
DR SMART; SM00298; CHROMO; 2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS50013; CHROMO_2; 1.
KM ATP-binding; Helicase; Hydrolase.
SQ SEQUENCE 5322 AA; 573615 MW; C9608375FA71C211 CRC64;
Query Match 5.4%; Score 95.5; DB 5; Length 5322;
Best Local Similarity 20.1%; Pred. No. 1.6e+02;
Matches 71; Conservative 49; Mismatches 139; Indels 95; Gaps 17;
QY 1 RVSGIGSRHHMPPYKRFPPDPDYCAQKTFCTGSPIPYMGDDDDIEVRLQAPW 60
DB 1993 RNFKIPQSEW-----KSKRRPPELWKKLEKT-----PYKGGNLRPYQLEGW 2039
QY 61 EPKGDILGHLXIMDAIGFSTLTGKNTYMEYELFQNGCTFPHLRPMADPFMONG 120
DB 2040 -LKFSVYNNHNCILDENGLGKTIGSLTFVSHVYEGING--PFLVIALPSTPMQR- 2095
QY 121 AACFFGIDIVHWKENGITLVQATISGNFNOMAKVKQDNFTGIYFTWYVASE-- 177
DB 2096 ---FEG-----WTDNNAVYVHGSV-----SKQIDYE--YYKTESGVLKEPIK 2137
QY 178 -KGAETWFSYDSCKVLRFTN-----KLAFFGAEFKNIE----- 211
DB 2138 ENVLITTFEMITVDYMDLKAFFWRLCVIDEARKLNKNCKLE-GIRQLNLRVLLSGT 2196
QY 212 ---TWYTRIF--LYSGEPYLYGNETSVEFGPTGNKTLGLAIKFFYFPKPHL----- 257
DB 2197 PLQNNISELFSLLNLEPQSFSSQEFMSFSLTKEEVNKLQALKKMLRLKLDYVE 2256
QY 258 ---PTKEFLLLQLQFDPAVIVHKQFYLFNFEYFLPKFPPIK--ITYEIP 305
DB 2257 KSLAPKEETILIV-----ELTNIQKKY-----RGLLEQNFSLKGTTSANIP 2300
RESULT 13
Q9TYM2 PRELIMINARY; PRT; 362 AA.
AC Q9TYM2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Y25CIA.7a protein.
GN Y25CIA.7 OR Y25CIA.7A.
OS Caenorhabditis elegans.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kerhaw J., Kirsten U., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders B., Showkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thirion-Mieg J., Thomas K., Vaudin N., Vaughan K., Waterston R.,
RA Watson A., Weinstein L., Wilkison-Sproat V., Wohlman P.,
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

elegans";
 Nature 368:32-38(1994).
 (2)
 SEQUENCE FROM N.A.
 STRAIN=Bristol N2;
 Kalicki J., Smith A., Gibson A.,
 "The sequence of C. elegans cosmid Y25C1A",
 Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
 (3)
 SEQUENCE FROM N.A.
 STRAIN=Bristol N2;
 Waterston R.;
 Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
 EMBL; AF125459; AAD12838.1; -
 PIR; T33904; T33904.
 Wormpep; Y25C1A.7a; CE21474.
 InterPro; IPR006977; DUF649.
 Pfam; PF04893; Y1P1.1.
 SEQUENCE 362 AA; 40613 MW; 42DBA5A7F72F3B89 CRC64;

Query Match 5.3%; Score 94; DB 5; Length 362;
 Best Local Similarity 21.7%; Pred. No. 64; Mismatches 88; Indels 58; Gaps 9;
 Matches 48; Conservative 27;

```

73 IMHDAIGRSTLTGKNTWMEYELF-----QLGNCCTFPHLR-----PEMD 112
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48 VEDSQSTSTSRKCNPFSEFYQQFEDVETDQVIRKLNLSVIFPHRNVIQDFLQIPIDLM 107
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
113 AEFNONGAACEFGIDVHAKENGTLVQVATISGNFMQAKVKKQDNETGIYETWNV 172
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
108 GPEVAVS-----TLVFALIGFNG-----LAQFIENDGAKGTGSDFRM 145
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
173 KASPEKGAET---TWEDSYDCS--KFLVTFNKLAERGAEFKNIENNYRIFLYSEPT 225
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
146 AKIENNTAISIFAQPPKKNCKTIIDSMAKGFQKLNLCSEFFKKS-----FCHQNSPS 200
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
226 YLGNSTVPGFTGKTLGLAIK---FYYPEKPHLPYKEFL 263
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 FC-MISHGFLTNGMLKTGSFSEQKMFYTSHTSASTLIFL 240

```

SUIT 14
 5330

025330 PRELIMINARY; PRT; 1238 AA.

01-JAN-1998 (TREMBlrel. 05, Created)
 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 Hypothetical protein HP0609.

HP0609.
 Helicobacter pylori (Campylobacter pylori).
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 Helicobacteriaceae; Helicobacter.
 NCBI_Taxid=210;

(1)
 SEQUENCE FROM N.A.
 STRAIN=26695 / ATCC 700392;
 MEDLINE=97394467; Pubmed=9252185;
 Tomb J.-F., White O., Keriavage A.R., Clayton R.A., Sutton G.G.,
 Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 Berg D.E., Gocayne J.D., Ullrich T.R., Peterson J.D., Kelley J.M.,
 Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
 Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 Venter U.C.;

"The complete genome sequence of the gastric pathogen Helicobacter
 pylori".
 Nature 388:539-547(1997).
 EMBL; AF000575; AAD07677.1; -
 PIR; A64596; A64596.
 TIGR; HP0609; -

DR GO; GO:0009279; C:external outer membrane (sensu Gram-negativ. . .; IEA.
 DR GO; GO:0015070; F:toxin activity; IEA.
 DR InterPro; IPR004311; Ptc_vacu_cyttox.
 DR Pfam; PF03077; VacA2; 3.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1238 AA; 135062 MW; 66DF754EB1BF8173 CRC64;

Query Match 5.3%; Score 94; DB 16; Length 1238;
 Best Local Similarity 23.6%; Pred. No. 32;
 Matches 62; Conservative 28; Mismatches 77; Indels 96; Gaps 15;

```

QY 80 FRSTLTGKNTWMEYELFQLGNCCTFPHLRPEMDAPFMCN-----QGAACFPFGIDVHWK 134
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 513 FTQTYGKNSALVF-----NATTP-----WANGSIPKSNSTVRGEGVWVG 555
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 135 ENGLVQVAT-----ISGNFMQAKVKKQDNETGIYETWNVASPEKGAETWPSYDC 189
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 556 KTYITGTFADRYVYIGNW-----IG-----NGAQTGGA--T 588
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 190 SKFLRTFNKLAERGAEFKNIET--NYTRIFLYSGEFTYIGNETSVFGPTGKTLGLAI 246
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 589 LNFVGAETENIA--GATFKMLKITSQNSYTMFMALG-----SSGSAKINVSQ 634
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 247 KRFY-----YPEKPHLPYKEFLSLQIFPAVIVHQFYLF-----YNEFYV-FLPMK 293
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 635 SDFYDWTGGGYDFYFGN-----GVFDSVFNFRAYVFGQTEMSYNKNTNFMALGN 683
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 294 PPT-KITTEIPIPIPKNTLSG 315
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 684 FKFGKTTIEKSVLSDASYTFDG 706
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

RESULT 15

08WHM9 PRELIMINARY; PRT; 2313 AA.

AC 08WHM9;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN YCF2.
 OS Psilotum nudum (Whisk fern).
 OC Chloroplast.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Moniliiformopses; Psilotophyta; Psilotales; Psilotaceae; Psilotum.
 OX NCBI_Taxid=3240;
 RN (1)
 SEQUENCE FROM N.A.

RC STRAIN=cv. Kingyoku;
 RA Makasugi T., Nishikawa A., Yamada K., Sugiyura M.;
 RP "Complete nucleotide sequence of the chloroplast genome from a fern,
 Psilotum nudum".
 RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RT EMBL; AP004638; BAB84296.1; -
 DR GO; GO:0009507; C:chloroplast; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0005351; F:sugar porter activity; IEA.
 DR GO; GO:0009401; F:phosphoenolpyruvate-dependent sugar phospho. . .; IEA.
 DR InterPro; IPR003959; AAA_ATPase_centre.
 DR InterPro; IPR002114; HPr_Serp_S.
 DR Pfam; PF00004; AAA_1.
 DR Pfam; PF05695; DUF825; 1.
 DR PROSITE; PS00589; PTS_HPR_SSR; 1.
 KW Hypothetical protein; Chloroplast.
 SQ SEQUENCE 2313 AA; 273575 MW; FE005DF98460C03B CRC64;

Query Match 5.3%; Score 94; DB 8; Length 2313;
 Best Local Similarity 21.7%; Pred. No. 72;
 Matches 62; Conservative 55; Mismatches 103; Indels 66; Gaps 16;

```

QY 79 GPRSTLTGKNTWMEYELFQLGNCCTFPHLRPEMDAPFMCNQAACFPFGIDVHWK 134
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

2b 1059 GRSFIRKRGFCLEGTNFFQIKSHLFNKI-PNKEYPH-----SITQTIQILHRFYLIK 1.11
2y 135 ENGTLVQVATISG---NMENQMAKXQDNFTGIYETWVKASPEKGAETW---PDSYD 188
2b 1112 EFGSSIQIKSISTEQVNLFDLQERFL--NSSIRKQILWIGVS-----DYWQPLDSDP 1.63
2y 189 CSKFVL---RTENKLAERGAEEFNKI-ETNYTRIFLYSGEFTYLGK-----ET 231
2b 1164 TNDPHLMNISTKQDOLNQGSGSIIDEKSYHNDYLS--KFGNLEBYDMLFRLKIPEL 1220
2y 232 SV-FGPTGNKTLGLA-----IKRFYYP--FKPHL---PTKEFLLSLQIFDAVIY 275
2b 1221 SIHETPDOSKISLESLEKHEFNQDIKNITYVDKNIFRSNLMEFNERLNVINTELLRVSTI 1280
2y 276 HKQFYLFYNEFYWFLPMKF-----PFIKITYEEIPLPIRNKTLISGL 316
2b 1281 AKKMLFPHEYIIPWFETIEMWKYINSAVNTPSETLNINISDQWISNL 1326

Search completed: June 8, 2004, 14:07:07
Job time : 59.3722 secs